



**Titre:** Modèle hybride d'estimation de la demande de transport collectif  
Title:

**Auteur:** Fanja Rajesson  
Author:

**Date:** 2009

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Rajesson, F. (2009). Modèle hybride d'estimation de la demande de transport collectif [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie.  
Citation: <https://publications.polymtl.ca/8444/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/8444/>  
PolyPublie URL:

**Directeurs de recherche:** Martin Trépanier, & Catherine Morency  
Advisors:

**Programme:** Génie industriel  
Program:

UNIVERSITÉ DE MONTRÉAL

MODÈLE HYBRIDE D'ESTIMATION DE LA DEMANDE DE TRANSPORT  
COLLECTIF

FANJA RAJESSON  
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(GÉNIE INDUSTRIEL)  
AOÛT 2009

© Fanja Rajesson, 2009.



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*  
ISBN: 978-0-494-53923-1  
*Our file Notre référence*  
ISBN: 978-0-494-53923-1

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé:

MODÈLE HYBRIDE D'ESTIMATION DE LA DEMANDE DE TRANSPORT  
COLLECTIF

Présenté par : RAJESSON Fanja

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. AGARD Bruno, Doct., président

M. TRÉPANIÉ Martin, ing., Ph. D., membre et directeur de recherche

Mme MORENCY Catherine, ing., Ph. D., membre et codirectrice

M. TRUDEAU Pierre, M. Sc., membre

## REMERCIEMENTS

Je tiens à témoigner toute ma reconnaissance envers mon directeur de recherche, Martin Trépanier, et ma co-directrice, Catherine Morency, pour leur dévouement et leur patience dans mon apprentissage et pour toutes les connaissances qu'ils m'ont inculquées durant cette maîtrise.

Je remercie également ma famille et mes proches qui m'ont soutenue durant toute la période d'études et de recherche et ce malgré leur éloignement.

Je remercie, enfin, mes collègues étudiants des départements de génie industriel et de génie civil qui m'ont permis de réaliser mon projet de recherche dans les meilleures conditions.

## RÉSUMÉ

La planification dans le domaine des transports est une étape fondamentale qui permet de déterminer efficacement les besoins de l'ensemble d'un système de transport: réseau et infrastructures, flotte (pour le transport public), usagers. L'objectif de la planification est de répondre à l'équilibre nécessaire entre l'offre et la demande sur un territoire donné.

L'estimation de la demande représente une part essentielle du travail du planificateur puisqu'elle permet d'adapter et de développer le réseau par rapport aux usagers et plus précisément par rapport à leur comportement. Les déplacements sont modélisés à partir de différentes techniques de collecte de données socio-économiques pertinentes (sondage) telles que les enquêtes ou recensements (selon différentes méthodes: téléphonique, internet, etc.) sur les déplacements, sur la population et les ménages et sur les activités socio-économiques, ainsi que les données du réseau (infrastructures, capacités des routes, lignes et arrêts du transport en commun).

Les objectifs actuels des transports visent l'efficacité du réseau (circulation fluide, accessibilité, sécurité, intégration de tous les modes), la préservation de l'environnement et la durabilité des modes c'est pourquoi le perfectionnement des outils de planification est primordial pour aider à l'évolution du système de transport.

Ainsi, la problématique du mémoire est de pouvoir effectuer l'estimation de la demande pour le cas du Transport en Commun (TC). Elle repose donc sur l'élaboration d'outils techniques et/ou théoriques qui permettraient de comprendre le comportement des voyageurs ainsi que de localiser et de quantifier les usages du TC. Afin d'appliquer notre étude à un cas réel, nous avons choisi de réaliser l'évaluation du réseau de transport de la ville de Montréal au Canada.

La planification spécifique au TC est indispensable puisqu'elle permet à l'organisme en charge d'adapter sa flotte et son réseau par rapport aux besoins de la population. La complexité de l'étude réside dans l'identification des caractéristiques des usagers, des déplacements mais aussi des activités socio-économiques du territoire susceptibles d'influencer la réalisation d'un déplacement en TC au détriment d'un autre mode de transport. D'ailleurs, nous allons voir que la majorité des théories touchant la modélisation d'un réseau de transport considère rarement le TC en mode seul. Les données de notre étude de cas nous permettent d'isoler ce mode afin d'appliquer aisément notre modèle. Le projet représente donc une innovation dans le domaine de la planification des transports.

Dans un premier temps, nous allons réaliser une revue historique de la recherche dans le domaine de la planification des transports. Depuis l'essor de l'automobile après la Seconde Guerre Mondiale, la mentalité et les objectifs des planificateurs et des gestionnaires ont bien évolué. Nous verrons comment, au cours des décennies, les problématiques ont été modifiées au point de vouloir changer le comportement des usagers. Les problèmes environnementaux et des limites de capacité des réseaux sont apparus et demandent non seulement une restructuration des réseaux mais aussi des pratiques.

En second, nous présenterons la Procédure Séquentielle Classique (PSC). Il s'agit du processus traditionnel à quatre étapes de l'estimation de la demande : Génération, Distribution, Affectation, Choix Modal. Nous y recenserons les méthodes les plus répandues, la méthodologie ainsi que ses limites. Nous avons ajouté la collecte de données à la procédure car il s'agit d'une étape à part entière dans la planification.

La suite du mémoire portera sur le projet de recherche en lui-même. L'objectif principal est de développer une méthode se basant sur la PSC et qui s'adapte à tout type de ville ou région pour estimer le nombre d'usagers qui effectuent un déplacement TC à l'intérieur du territoire. À partir des données à disposition, au minimum des données de recensement et

de référence sur l'usage du TC, le processus de modélisation doit produire une ou plusieurs matrices Origine-Destination (OD) représentant la demande de déplacements en TC entre les différentes zones d'origine et de destination appartenant au territoire.

L'adaptation de la PSC appliquée au TC comporte trois étapes : la collecte et le traitement des données, la génération des déplacements et la distribution (création des matrices OD). Les objectifs et la démarche de cette méthode hybride sont expliqués de façon théorique avant d'être appliqués au cas réel de l'île de Montréal. Cette alternative doit permettre d'obtenir de meilleurs résultats qu'avec la méthode classique avec les outils à disposition. Pour fins d'expérimentation, le projet s'appuie sur l'utilisation quasi-exclusive du logiciel Microsoft Excel pour réaliser l'ensemble des étapes de la méthode hybride. Cette contrainte permet de cadrer l'étude mais surtout diminue les capacités de résolution du problème d'estimation.

Le traitement des données adapte les données en fonction du découpage territorial désiré. Dans le cas de Montréal, le système zonal choisi correspond aux municipalités et se compose de 40 zones. L'analyse des données socio-économiques, de population, de réseau TC et de déplacements qui ont été recensées pour chacune des zones permet de définir des classes de population par rapport aux déplacements TC. Par exemple, les usagers de 15-29 ans, ceux de 30-44 ans et ceux de plus de 65 ans n'utilisent pas le TC pour les mêmes raisons et ne présentent pas le même niveau de demande. Les zones dont le taux de chômage est supérieur à 8% sont moins enclines à emprunter le TC par rapport aux zones où le taux est inférieur.

La génération des déplacements estime les coefficients déterminant les nombres de déplacements entrant et sortant de chaque zone par rapport aux variables de l'étude (socio-économiques, de population et de réseau TC). Le premier constat indique que le nombre de passages du métro en période de pointe est la variable la plus pertinente pour expliquer les



déplacements à partir de et arrivant dans chaque zone. De plus, les modèles sont performants puisqu'ils réussissent à expliquer plus de 80% de la variabilité des observations sur l'ensemble du territoire pour le découpage municipal. L'affinage des résultats pour le Centre Ville (CV) donne de meilleurs résultats comparativement à la méthode classique : les résultats obtenus pour la majorité des zones ne présentent pas plus de 30% de décalage avec la réalité (en nombre de déplacements) contrairement à ceux de la PSC qui présentent un nombre non négligeable de zones donc le décalage est supérieur à 70%.

L'étape de distribution met en évidence les limites de l'outil notamment par rapport à la capacité du solveur Excel (en nombre de contraintes). Ainsi, la division en sous-matrices permet de pallier aux lacunes et d'obtenir des résultats plus ou moins intéressants. Par ailleurs, l'analyse des corrélations entre la matrice OD obtenue et la matrice de référence révèle le manque de données pertinentes intégrées dans l'évaluation des coefficients qui déterminent la propension d'un usager à se déplacer à partir ou à destination d'une zone. In fine, la matrice présente des résultats intéressants qui sont très bien exploitables dans certains cas de déplacements. Le profil type de l'utilisateur du TC peut être défini pour chaque zone d'origine ou de destination.

Le modèle obtenu s'apparente à la procédure classique, aux travers des étapes de génération et de distribution, tout en délivrant de meilleurs résultats dans l'étude de cas. La particularité de la démarche (d'où la méthode hybride) se caractérise par une approche progressive, par échelle de découpage et par type de destination des déplacements, et par la validation des résultats à partir de leur interprétation par rapport aux données du territoire et non seulement à partir d'une matrice de référence. Elle propose de nouvelles lignes de recherche dans le domaine des transports en reliant intimement la population et les activités socio-économiques (la demande) au système de transport (l'offre) en se basant essentiellement sur l'interprétation des comportements types de l'utilisateur.

## ABSTRACT

Transport planning is a fundamental element of an efficient transport system management. Estimating and forecasting the needs in network and infrastructure, equipment (fleet, signals) and demand (users) depending on the actual system give guidelines to keep equilibrium between transport supply and demand.

Transport demand estimation represents an important part of the planner work allowing the transport system to be adapted and developed to meet the trips need through the study of the user trips and more precisely their behaviour. Trips can be modelled from different data collection techniques: surveys and census on population (age, activity, mobility, travels), households and socio-economic activities; and from data describing the transport network (infrastructure, capacity, routes, accessibility, transit stops and stations).

The current objectives of transport focus on the network efficiency (smooth traffic, accessibility, security, multimodal transport), environment protection and sustainable growing. Consequently, improvements of the planning tools to help meeting these objectives are essential.

The subject of the memoir deals with the demand estimation for the Transit mode. The research is based on the creation of technique and/or conceptual tools which would help to understand the user behaviour through the interpretation his or her profile and Transit use (localisation and quantification of trips). It also includes experimental applications for the case of the city of Montreal in Canada.

Transit planning is essential as it allows the transport agency to adapt its fleet and network geometry to the population needs. The complexity of the research lies in the identification of relevant user characteristics, trips and socio-economic activities of the region likely to influence any transit trip. Besides, the majority of transport network modelling theories

never focus only on transit but on multiple modes at the same time. The data used for the experiments represents this specific mode so that our tools can easily be applied. Then, this research can be considered as an innovation in transport planning.

First and foremost, an historic and literacy review of the transport planning domain is made. From the rapid development of the automobile after the World War II, the mentality and objectives of transport planners and managers has completely evolved. During decades, the transport problems have been modified to the point of changing user behaviour. Environmental issues and network capacity limits were reached and imply the redevelopment not only of the network but also the practices.

Secondly, the sequential four-step transportation forecasting process is introduced. Divided into four steps: Generation, Distribution, Mode Choice and Route Assignment, this procedure estimates the number of travelers that will use a specific transportation system. The most used methods and their limits are presented. The data collection (which precedes the process) is inserted as a step as this procedure should be formalized for all the possible case study.

The following chapters deal with the research project itself. The principal target is to develop a method based on the sequential process which can be adapted to any city or region to estimate the number of transit trips made inside the territory. From the collected data (minimum of census and essential transit data), the process has to produce Origin-Destination matrices representing the transit demand in number of trips for each pair of origin and destination zone of the territory.

The adaptation of the sequential process applied to transit consists of three steps: data collection and processing, trip generation and trip distribution (OD matrices creation). Objectives and methods of this hybrid method are conceptually explained before the

experimental application to Montreal. This alternative has to give better results than with the classical method with the same tools which are at disposal. For the research, the experiments are made with the quasi-exclusive use of Microsoft Excel. This constraint gives a frame to the study but reduce the solving capacities of the problem.

Data processing adapts all the data collected to the given territory zoning system. For our case study, this system corresponds to the municipal districts, which represents 40 zones. The analysis of socio-economic, population, transit data and effective transit trips allows the identification of population classes depending on transit trips. For instance, age groups of 15-29, 30-44 and more than 65 users do not use transit for the same reasons and at the same demand rate. People from zones where the unemployment rate is higher than 8% are less likely to use transit as a mode of transportation than the others.

Trip generation estimates coefficients which determine the number of trips entering and going out of each zone depending on the socio-economic, population and transit variables. The first report indicates that the metro traffic during the peak period is the most relevant variable explaining such trips. In addition, the model gives excellent results with more than 80% of the variability explained. The refining of the results for the city center gives better results than the sequential process: the majority of the results are less than 30% different from reality compared to 70% for the old method.

The distribution step shows the limits of the tool particularly for the Excel solver. The segmentation of a matrix into different sub-matrices compensates for the capacity limits and gives much or less better results. Moreover, the correlation analysis between the trip estimations and the reference data reveals the lack of relevant variables determining the transit user behaviour from an origin or to a destination. In fine, the OD matrix gives relevant results which can be used to explain different transit trip trends in the region. The typical user profile can be defined for each zone as an origin or a destination.

The resulting model is similar to the sequential process, through the generation and distribution steps, giving overall better results in the case study. The particularity of this hybrid process resides in the progressive approach, by zoning system scale and trip destination type, and the results validation by their explainability through population, socio-economic, transit data. The memoir suggests research opportunities for the domain intimately relating the population and its socio-economic activities (demand) to the transportation system (supply) based on the interpretation of the typical user behaviour.

## TABLE DES MATIÈRES

REMERCIEMENTS .....	IV
RÉSUMÉ .....	V
ABSTRACT .....	IX
TABLE DES MATIÈRES .....	XIII
LISTE DES TABLEAUX.....	XVII
LISTE DES FIGURES.....	XVIII
LISTE DES ANNEXES.....	XX
CHAPITRE 1. INTRODUCTION .....	1
1.1 Problématique du projet .....	1
1.2 Objectifs du mémoire de recherche.....	2
1.3 Méthodologie et structure du mémoire .....	2
CHAPITRE 2. REVUE DE LITTÉRATURE .....	5
2.1 Bref historique jusqu'au milieu du 20 <sup>e</sup> siècle.....	5
2.2 La planification dans le domaine des transports urbains jusqu'en 1990.....	8
2.3 L'état de la recherche en planification .....	15
2.4 Synthèse .....	27
CHAPITRE 3. LA PROCÉDURE SÉQUENTIELLE CLASSIQUE .....	28
3.1 Collecte des données .....	29
3.2 Génération.....	31
3.2.1 Le système de zonage : Agrégation-Désagrégation.....	32
3.2.2 Méthode la plus utilisée : la Régression Linéaire .....	32
3.2.3 Régression par classe (ou par catégorie).....	34
3.2.4 Les indicateurs de performance .....	35
3.2.5 Critique.....	35
3.3 Distribution .....	37
3.3.1 Le modèle à facteurs de croissance.....	38

3.3.2	Le modèle gravitaire .....	39
3.3.3	Le modèle d'opportunité .....	40
3.3.4	Les indicateurs de performance .....	41
3.3.5	Critique.....	41
3.4	Répartition Modale.....	42
3.5	Affectation.....	43
3.6	Outils de planification .....	43
3.6.1	GNE7/QRSII (AJH Associates).....	43
3.6.2	TLOS.....	44
3.6.3	TBEST .....	47
3.7	Synthèse .....	48
CHAPITRE 4. MÉTHODOLOGIE.....		49
4.1	Outils du projet.....	50
4.1.1	MICROSOFT OFFICE EXCEL .....	50
4.1.2	FGIS (© Université 2003 du Wisconsin).....	51
4.1.3	STATA (STATACORP LP) .....	51
4.1.4	MICROSOFT MAPPOINT .....	51
4.2	Collecte et traitement des données.....	52
4.2.1	Définition du système de zonage : agrégation-désagrégation.....	52
4.2.2	Analyse préliminaire des variables .....	53
4.2.3	Stratification des variables .....	54
4.3	Génération des déplacements.....	55
4.4	Distribution des déplacements .....	58
4.5	Synthèse .....	62
CHAPITRE 5. TRAITEMENT DES DONNÉES.....		64
5.1	Sources de données .....	64
5.1.1	Territoire .....	64
5.1.2	Recensement .....	64

5.1.3	Caractéristiques du réseau.....	66
5.1.4	Enquêtes origine-destination.....	67
5.2	Structures de données proposées.....	68
5.2.1	Données territoriales .....	68
5.2.2	Données opérationnelles .....	68
5.2.3	Données sur les paires origine-destination.....	68
5.3	Choix du système zonal .....	69
5.4	Analyse préliminaire : classification des variables .....	71
5.4.1	Premières Transformations des variables.....	71
5.4.2	Étude des déplacements par rapport aux données de population.....	74
5.4.3	Étude des déplacements vers Centre Ville/Hors Centre Ville .....	78
5.4.4	Stratification des variables .....	79
5.5	Synthèse .....	83
CHAPITRE 6. GÉNÉRATION DES DÉPLACEMENTS.....		84
6.1	Étape 1 : Génération classique.....	84
6.1.1	Méthodes de sélection des variables .....	85
6.1.2	Génération des déplacements à partir d'un solveur .....	86
6.1.3	Quelques résultats d'expérimentation.....	87
6.2	Étape 2 : Génération/Distribution .....	92
6.2.1	Quelques résultats d'expérimentation.....	93
6.3	Synthèse .....	96
CHAPITRE 7. DISTRIBUTION DES DÉPLACEMENTS.....		97
7.1	Première distribution : le modèle le plus agrégé.....	97
7.1.1	Détermination de la fonction d'impédance .....	97
7.1.2	Modèle gravitaire .....	97
7.1.3	Calage du modèle.....	99
7.1.4	Modèle gravitaire en utilisant le solveur .....	99
7.1.5	Indicateurs de reproductibilité .....	100



7.1.6 Quelques résultats d'expérimentation .....	100
7.2 Distribution par désagrégation : affinage .....	103
7.2.1 Quelques résultats d'expérimentation .....	104
7.3 Validation de la matrice complète .....	106
7.4 Synthèse .....	108
CONCLUSION .....	110
Analyse critique du modèle classique .....	110
Analyse critique du modèle hybride .....	111
Synthèse .....	115
BIBLIOGRAPHIE .....	117
ANNEXES .....	120

## LISTE DES TABLEAUX

Tableau 2.1: Récapitulatif des avancées de la recherche dans le domaine des transports jusqu'en 1950.....	8
Tableau 2.2: Récapitulatif des avancées en planification dans le domaine des transports urbains jusqu'en 1990 .....	14
Tableau 2.3: Récapitulatif des avancées récentes en planification dans le domaine des transports urbains .....	27
Tableau 5.1: Variables de déplacements à modéliser .....	73
Tableau 5.2: Variables explicatives de l'étude .....	82
Tableau 7.1: Comparaison des variables pertinentes expliquant la distribution des déplacements attirés pour les secteurs de destination 101, 102 et 127 .....	102
Tableau 7.2: Matrice de corrélations entre les déplacements à destination de certaines zones du CV et les données de population et du réseau pour l'île de Montréal .....	108
Tableau A.1: Comparaison du nombre de déplacements entrants à l'intérieur des zones du CV, par motif de déplacement .....	120
Tableau A.2: Comparaison du nombre de déplacements entrants à l'intérieur des zones hors CV, par motif de déplacement .....	122
Tableau B.1: Exemple de matrice OD pour les secteurs municipaux 101 à 112 du cas de Montréal.....	125

## LISTE DES FIGURES

Figure 1.1: Déroulement du mémoire de recherche.....	4
Figure 2.1: L'explosion du secteur Automobile (Motor Vehicle Registrations, Fuel Consumption, User Taxes and Highway Expenditures, 1910-1955) - U.S. Dept. of Commerce, 1954 .....	9
Figure 2.2: Modifications apportées au processus de distribution par Bonnel, 2004 .....	19
Figure 2.3: Modèle d'estimation de la demande de Bernardin (Bernardin, 2008).....	22
Figure 3.1: Procédure Séquentielle Classique.....	29
Figure 3.2: Matrice Origine-Destination avec les contraintes aux marges .....	37
Figure 3.3: Exemple d'application simple du logiciel TLOS issu du guide de l'utilisateur (Kittelton & Associates, Inc., 2001) .....	45
Figure 3.4: Exemple de résultats obtenus par TLOS pour une région urbaine en pointe du midi (Kittelton & Associates, Inc., 2001).....	46
Figure 4.1: Méthode hybride d'estimation de la demande TC .....	50
Figure 4.2: Première étape de la PSC - Génération de la demande .....	56
Figure 4.3: Première étape de génération des déplacements.....	56
Figure 4.4: Seconde étape de génération des déplacements .....	57
Figure 4.5: Comparaison des méthodes de génération classique et hybride.....	58
Figure 4.6: Matrice de distribution des déplacements .....	59
Figure 4.7: Distribution des déplacements pour la méthode hybride.....	60
Figure 4.8: Comparaison des méthodes classique et hybride de distribution des déplacements .....	62
Figure 5.1: Découpage territoriale en secteurs municipaux de l'Ile de Montréal .....	70
Figure 5.2: Découpage secondaire du Centre Ville de Montréal (SM101 et SM102).....	70
Figure 5.3: Nombre de déplacements entrants par SR.....	75
Figure 5.4: Nombre de déplacements des usagers par rapport à leur sexe (homme:1, femme:2) et leur tranche d'âge .....	76
Figure 5.5: Répartition des déplacements entrants ayant pour motif le travail.....	77

Figure 5.6: Répartition des déplacements entrants ayant pour motif les études .....	77
Figure 5.7: Répartition des déplacements à destination Hors-CV de Montréal.....	79
Figure 5.8: Définition des groupes d'âge (selon les cohortes en abscisse) par rapport au nombre de déplacements (en ordonnée).....	80
Figure 5.9: Stratification de la variable taux d'activité par rapport au taux de déplacements .....	81
Figure 6.1: Résultat de la génération des déplacements sortants pour le découpage principal de l'île de Montréal à partir de STATA avec application des méthodes de sélection de variables .....	88
Figure 6.2: Comparaison des déplacements sortants réels et ceux estimés par le solveur EXCEL.....	89
Figure 6.3: Résultats de la génération des déplacements attirés pour le découpage principal de l'île de Montréal à partir de STATA avec application des méthodes de sélection de variables .....	90
Figure 6.4: Comparaison des déplacements attirés réels et estimés par le solveur EXCEL	92
Figure 6.5: Étape de génération-distribution, comparaison en pourcentage des déplacements produits ( $E_i$ ) réels par rapport à ceux estimés par la méthode hybride .....	94
Figure 6.6: Étape de génération-distribution, comparaison en pourcentage des déplacements émis ( $E_i$ ) réels par rapport à ceux estimés par la PSC .....	94
Figure 6.7: Comparaison des déplacements produits réels et estimés par génération classique .....	95
Figure 7.1: Comparaison des déplacements produits se destinant à la zone 101 estimés avec ceux de la matrice de référence .....	103
Figure 7.2: Comparaison en pourcentage des déplacements entrants dans la zone "4620062" estimés par rapport à la matrice de référence .....	105
Figure 7.3: Comparaison en pourcentage des déplacements entrants dans la zone "4620062" estimés par rapport à la matrice de référence.....	106

## **LISTE DES ANNEXES**

ANNEXE A – Étude des déplacements par motif pour le cas de la ville de Montréal.....120

ANNEXE B – Exemple de matrice OD.....125

## CHAPITRE 1. INTRODUCTION

La planification des transports est le processus méthodologique et dynamique qui gère l'ensemble des composantes associées à la mobilité de la population, des biens et des services d'un espace géographique déterminé (Kanafani, 1983). Il permet à un système de transport de s'intégrer harmonieusement et de favoriser l'activité de la société. L'accessibilité au transport et la qualité des déplacements (conditions du transport, fiabilité des horaires, efficacité du réseau) sont les objectifs premiers de la planification.

Avant les années 80, le planificateur avait un rôle essentiellement d'analyste technique. Il fallait accompagner la croissance économique et donc le besoin de se déplacer dans les pays riches; la seule solution : construire des routes. La logique s'est alors renversée face aux problèmes majeurs engendrés par l'expansion du réseau routier tels que la congestion, les effets sur l'environnement et les édifices historiques ou la discrimination (contre les pauvres). Aujourd'hui, le planificateur doit intégrer de nouvelles notions notamment celle de développement durable (pour l'environnement, les modes de transport, l'aménagement du territoire, l'économie) et de management du trafic.

L'estimation de la demande a toujours été une étape primordiale du processus. À partir de l'étude du comportement des usagers et de l'aménagement du territoire, le planificateur détermine le nombre de déplacements effectués dans un espace donné qui est divisé en zones. La comparaison des résultats du modèle et de l'offre du système de transport en place permet d'évaluer les performances du réseau et/ou de définir concrètement la structure d'un réseau idéal.

### 1.1 Problématique du projet

La problématique du mémoire est de pouvoir effectuer l'estimation de la demande en termes de nombre de déplacements en transport en commun (TC). À partir de données **démographiques** et socio-économiques génériques, le projet de recherche consiste à mettre

en place une démarche scientifique qui implique des outils de traitement de données, de statistiques et de résolution de problèmes sous contraintes qui permettent d'expliquer le comportement, de localiser et de quantifier les usagers du TC. Ces outils doivent être facilement transférables d'un terrain d'étude particulier à un autre.

La ligne directrice du projet consiste à s'inspirer des premières étapes de la procédure séquentielle classique (PSC) qui constitue la base de la modélisation des déplacements. Il faudra alors adapter cette méthodologie au cas particulier du TC et l'améliorer.

Afin d'appliquer notre étude à un cas réel, nous avons choisi de réaliser l'évaluation du réseau de transport de la ville de Montréal au Canada.

## **1.2 Objectifs du mémoire de recherche**

Le projet vise à proposer un modèle d'estimation de la demande permettant, à terme, de construire la matrice des déplacements associée à un réseau de transport collectif.

Parmi les directions de recherche, il est visé entre autres de se distinguer de l'approche classique à quatre étapes, en tirant profit au maximum des données disponibles. De plus, une approche de génération-distribution hybride des déplacements est envisagée. Enfin, seuls les déplacements de transport collectif sont générés, ce qui exclut a priori les questions de transfert modal. Ainsi, il est supposé que la demande de transport collectif est en grande partie induite par l'accessibilité au réseau et les conditions sociodémographiques de la population.

## **1.3 Méthodologie et structure du mémoire**

Le modèle s'appuie donc sur trois étapes principales:

- Le traitement des données définit les variables qui seront utilisées pour estimer le modèle. L'analyse des chiffres collectés permettent d'extraire des variables d'étude (les variables explicatives du modèle) plus adaptées à la modélisation de la demande

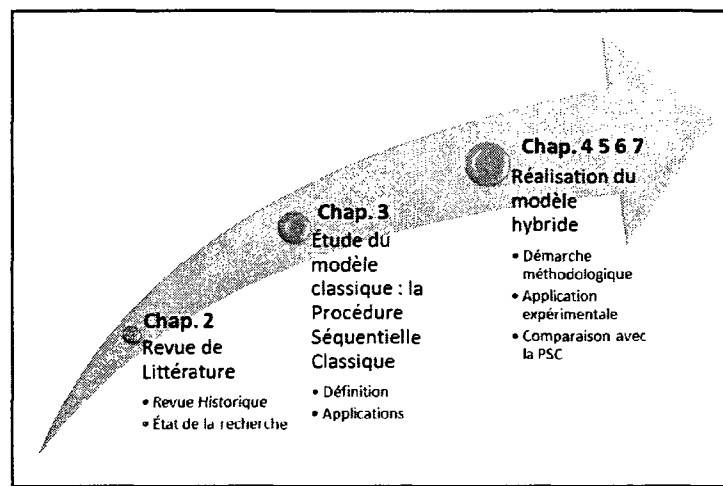
en transport. Ces (nouvelles) variables sont soit continues soit catégoriques et elles permettent de dresser un profil caractéristique pour chaque zone du territoire considéré.

- La génération de déplacements consiste à estimer le nombre de déplacements émis et attirés par chacune des zones. Comme intrants du cas de Montréal, des données traitées (étape précédente) issues du recensement et du réseau TC sont utilisés. La génération s'appuiera ainsi sur le service (offre) présent dans chacune des zones, une distinction par rapport aux modèles actuels.
- La distribution des déplacements consiste à déterminer les paires OD entre les zones, en considérant la demande générée.

La génération-distribution hybride consiste à réaliser les estimations par échelle de découpage et de distinguer les tendances types de déplacement. Cela s'applique très bien aux déplacements destinés au centre-ville de l'agglomération, car ceux-ci sont réguliers, pendulaires et assez bien circonscrits pour chacune des zones. Cette méthode nouvelle constitue également une contribution distinctive de l'approche classique (PSC).

Le mémoire s'articule ainsi en 6 chapitres retraçant la démarche du projet de recherche qui est décrite comme suit.





**Figure 1.1: Déroulement du mémoire de recherche**

Les deux premiers chapitres sont consacrés au constat de l'existant au travers des littératures et des avancées actuelles de la recherche du domaine des transports et à la présentation de la PSC, méthode à partir de laquelle a été développée notre méthode. Le troisième chapitre présente les idées porteuses du modèle hybride dont les étapes sont exposées de façon détaillées (formules, application expérimentale) dans les trois derniers chapitres.

## **CHAPITRE 2. REVUE DE LITTÉRATURE**

Cette revue de littérature nous permet de nous situer au niveau historique afin de comprendre la motivation du projet de recherche. L'évolution des enjeux socio-économiques et des moyens de transport ont influencé la recherche sur le transport et ont montré la nécessité des déplacements pour le développement de la société. Nous allons expliquer, depuis le début du 19<sup>e</sup> siècle ou l'essor de l'automobile, comment le transport est devenu une partie essentielle à la vie de tout un chacun et comment les chercheurs ont favorisés l'évolution du domaine.

### **2.1 Bref historique jusqu'au milieu du 20e siècle**

La mobilité est la colonne vertébrale de la société moderne. Pour chaque activité socio-économique, l'homme doit aller d'un endroit à l'autre et, pour se déplacer correctement, le moyen de transport emprunté doit être adéquat. La diversité des activités et la complexité des motifs définissent une infinité de déterminants au déplacement: du travail au loisir personnel. Ainsi, le transport est la composante principale des structures (économiques, sociales, culturelles, politiques) de notre société et le développement économique d'un secteur géographique – pays, région, ville, quartier – est complètement influencé par le transport au travers de la notion d'accessibilité (Kanafani, 1983).

L'adaptation du transport aux différents besoins de se déplacer a tout d'abord engendré la notion d'offre et de demande en transport. Kohl a étudié les relations entre la géographie des ressources et la structure des réseaux de transport. Il a mis en évidence l'interaction offre-demande par rapport à l'impact des moyens techniques de transport sur le trafic (Kohl, 1841,1850).

En effet, les premières études en transport ont soulevé le besoin essentiel d'analyse formelle et systématique de la demande pour pouvoir les comprendre et gérer les

déplacements. Kanafani définit l'analyse de la demande comme le processus qui met en relation la demande et les activités socioéconomiques qui la génèrent. Il estime que le nombre de migrants ( $I_{ij}$ ) entre un lieu campagnard  $i$  et une ville industrielle anglaise  $j$  est fonction directe de la population de la ville  $j$  et inverse de la distance séparant  $i$  et  $j$  ( $I_{ij} = P_j / d_{ij}$ ) (Kanafani, 1983).

Les interactions sont donc conceptualisées sous forme de modèles. Les premiers modèles sont apparus dès la fin du 19<sup>e</sup> siècle notamment avec Ravenstein qui a élaboré un modèle gravitaire pour l'étude des motifs de migration entre les villes (Ravenstein, 1885, 1889). Puis, Lill (1891) propose un modèle gravitaire pour analyser la planification d'un réseau de rails. Il définit la 'Loi du Voyage' (the law of travel) où il explique l'interaction entre deux points distincts de l'espace : "it makes no difference in the matter of time consumption and travel costs if a person with wanderlust makes  $n$  short trips of  $k$  kilometers or if the person makes one journey of the same length  $nk$ , as the wanderlust in both cases will be satisfied with the value of  $nk$ " (Lill, 1891, traduction 1969).

Durant la première moitié du XX<sup>e</sup> siècle, un certain nombre de sociologues et de démographes ont contribué au développement de l'analyse de la demande. Nous avons recensé les plus importants dont les travaux sont encore utilisés comme base aujourd'hui. Tout d'abord, inspiré de la loi de gravitation universelle de Newton, Reilly crée une loi qui établit que la puissance d'attraction d'une ville dépend de deux facteurs : l'inverse du carré de la distance et la population (de manière directe) (Reilly, 1931). Une personne serait prête à parcourir une plus grande distance pour atteindre une grande ville (plus peuplée). Cette loi de Reilly représente le premier pas rendant compte de l'attraction des aires urbaines. Une formulation du modèle de gravitation d'interaction qui spécifie la direction du déplacement est établie par Stouffer en 1940. Il essaie aussi de redéfinir la distance entre l'origine et la destination en fonction de la taille cumulative des destinations intermédiaires

(Stouffer, 1940). Il est le précurseur du modèle d'opportunités (opportunité de satisfaire un besoin au plus proche du lieu d'origine).

En 1947, Stewart introduit le concept de gravitation démographique (demographic gravitation) au travers de transpositions des comportements démographiques de la population (donc macroscopiques) à des notions classiques de la physique. A partir des hypothèses de la loi de Reilly, il a estimé que l'attractivité d'une zone est équivalente à la somme des attractivités des lieux isolés (points d'intérêt) appartenant à la zone. Par exemple, pour les Etats-Unis, le potentiel le plus grand se situe aux environs de New York et diminue graduellement comme des cercles concentriques autour de la zone sauf au niveau des aires urbaines où l'attractivité augmente (Stewart, 1948).

Zipf (1948) a examiné le phénomène d'interaction entre deux villes  $i$  et  $j$  pour des déplacements en transport en commun ou en avion ou même pour des appels téléphoniques par rapport au facteur «  $P_i P_j / D_{ij}$  » (où  $P$  est la population et  $D$  la distance). Il en a déduit que ce facteur élevé à une certaine puissance induisait une relation linéaire entre la variable d'étude et le facteur (Zipf, 1949).

Enfin, il est important de mentionner la formulation du « problème de transport » en programmation linéaire par Hitchcock (1941). Il s'agit d'un modèle simpliste qui minimise le coût global de transport pour un réseau de livraison devant satisfaire des contraintes de demandes et de capacités. Il s'agit d'une analogie avec le problème de production (« Industrial Cost Curve ») qui optimise la production d'un réseau d'usines par rapport au marché à approvisionner (Hitchcock, 1941).

Ces premières approches de modélisation nous montrent que les premiers objectifs du planificateur était de construire un réseau qui corresponde aux lignes de désir de la population. En fonction de l'importance, de l'attractivité d'une aire urbaine, il fallait

favoriser les activités socio-économiques. Cependant, ils n'ont pas pris en compte les phénomènes de congestion, de pollution ou d'étalement urbain qui poseront plus tard des problèmes dans la performance des déplacements.

Le tableau ci-après résume les différentes approches exposées qui représentent les bases de la recherche en transport.

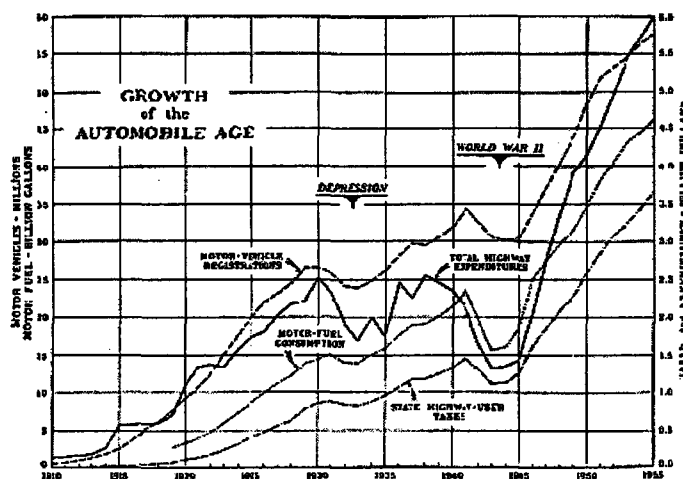
**Tableau 2.1: Récapitulatif des avancées de la recherche dans le domaine des transports jusqu'en 1950**

Année	Chercheur	Avancée
1841	Kohl	Interaction offre-demande
1885	Ravenstein	Etude des migrations entre les villes
1891	Lill	La Loi du Voyage
1931	Reilly	Modèle gravitaire inspiré de la loi de Newton
1940	Stouffer	Modèle d'opportunité
1941	Hitchcock	Programmation linéaire : le problème de transport
1947	Stewart	Concept de gravitation démographique
1948	Zipf	Phénomène d'interaction entre deux villes

## **2.2 La planification dans le domaine des transports urbains jusqu'en 1990**

La compréhension des changements de notre société est un élément vital pour l'homme qui veut façonner son environnement intelligemment. Il doit connaître la manière dont la population utilise le sol : les activités socio-économiques, les structures associées à chaque activité, leur répartition spatiale sans oublier les routes pour y accéder. A l'issue de la Seconde Guerre mondiale, les États-Unis veulent organiser le système de transport pour favoriser le développement socio-économique.

Les premiers modèles que nous avons énoncés précédemment ont essentiellement servi au niveau commercial (notion de géomarketing). En effet, ces études ont eu pour but de décrire le comportement d'un consommateur et de son besoin par rapport aux points d'attraction (magasins, points de vente). Le modèle gravitaire de Newton a montré que tout comme deux corps en présence d'une force suivent la loi physique d'attraction (ou de répulsion), l'être humain a un comportement similaire lorsqu'il choisit de se déplacer à un endroit au détriment d'un autre.



**Figure 2.1: L'explosion du secteur Automobile (Motor Vehicle Registrations, Fuel Consumption, User Taxes and Highway Expenditures, 1910-1955) - U.S. Dept. of Commerce, 1954**

Les études sur la planification du transport urbain des années 50-60 ont été une contribution analytique majeure. C'est dans cette période, où la voiture était utilisée comme moyen de déplacement essentiel et fiable (figure 2.1), que le rôle du planificateur a été d'adapter le réseau de transport à la croissance économique et donc à la demande. Auparavant, l'objectif du planificateur était de réaliser un réseau qui couvre l'ensemble d'un territoire : notion d'accessibilité. L'objectif ayant été atteint, le problème de l'augmentation du trafic se pose. En 1950, le U.S. Highway Research Board a publié un recueil de pratiques pour identifier les lignes de désir et relier les paires Origine-Destination (Campbell, 1950). C'est

effectivement au milieu des années 40 que les premières enquêtes OD sont réalisées aux Etats-Unis.

Le principe d'équilibre de Wardrop (1952) a permis de formaliser l'affectation des déplacements des usagers en considérant que ceux-ci tendent à emprunter les itinéraires les moins coûteux. Lorsqu'une congestion se produit sur sa route, l'usager va prendre un chemin alternatif qui puisse le mener à sa destination en un temps minimal : « *The journey times in all routes actually used are equal and less than those which would be experienced by a single vehicle on any unused route* » (Wardrop, 1952). De plus, il postule qu'au bout d'un certain temps, le temps de parcours moyen est minimal c'est-à-dire qu'il existe une forme d'équilibre dans le réseau où le comportement des usagers rend les parcours optimaux : « *At equilibrium the average journey time is minimum* ». C'est en 1955 qu'un modèle mathématique associé est réellement formulé au travers des travaux de Beckmann, McGuire et Winsten qui ont mené à la création d'un paradigme micro-économique de l'offre et de la demande incluant le transport : « *Demand refers to trips and capacity refers to flows on roads. The connecting link is found in the distribution of trips over the network according to the principle that traffic follows shortest routes in terms of average cost. The idea of equilibrium in a network can then be described as follows: . . . the existing traffic conditions are such as to call forth the demand that will sustain the flows that create these conditions* » (Beckmann, Mc Guire, & Winsten, 1956).

Afin d'améliorer la compréhension de la relation entre déplacement et utilisation du sol, Mitchell et Rapkin (1954) ont élaboré des méthodes empiriques pour les mouvements de personnes et de biens. Ils ont mis en évidence le fait que les déplacements partant du lieu de résidence varient en fonction du sexe et de l'âge des individus du ménage considéré. Le profil de l'utilisateur modifie significativement son comportement : les célibataires, jeunes mariés, familles avec enfants et personnes âgées n'ont pas les mêmes intérêts dans leurs déplacements (motif, destination) (Mitchell & Rapkin, 1954). C'est pourquoi les enquêtes

recensant tous les motifs de voyage et leur correspondance avec les habitudes des usagers et leurs déplacements permettent de comprendre leur comportement et d'augmenter le pouvoir de prédiction des modèles utilisés en planification. Oi et Shuldiner (1962) l'ont d'ailleurs expliqué dans leur étude : *“urban travel is treated as an aspect of human behavior, related in many ways to the over-all economic and social behavior of urban residents”* (Oi & Shuldiner, 1962). Ils ont étudiés les déterminants associés à chaque motif de déplacement (travail, étude, loisirs...) ainsi que leurs impacts à partir des données d'enquêtes OD.

Wilson reprend le modèle gravitaire introduit par Reilly et introduit la notion d'entropie dans la distribution spatiale des déplacements. Il a développé une procédure de maximisation de l'entropie qui est utilisée pour décrire l'allocation de la distribution des déplacements par rapport au comportement d'individus pour une population (Wilson, 1974).

En 1973, Hägerstrand introduit un modèle des déplacements dans le contexte spatio-temporel. Il y met en relation les activités humaines sous différentes contraintes de l'espace-temps et introduit la notion de « *time geography* » (Hägerstrand, 1952). Un exemple de contrainte serait des horaires de début et fin d'activités dispersées dans l'espace ; cette dispersion impose des déplacements qui eux-mêmes dépendent de l'offre et des moyens de transport dont dispose l'individu. L'apparition de cette troisième dimension permettrait aux planificateurs d'avoir une vision plus précise de la demande par rapport aux activités des usagers dans le temps.

Cependant, les travaux de Dupuy (1975) et Margail (1989) mettent en évidence l'influence des méthodes américaines de planification et de programmation des infrastructures de transports urbains par rapport aux pratiques des ingénieurs de transports français. Ils concluent que les premières méthodes ont produit une urbanisation favorisant la croissance de l'automobile et tendent à renforcer les pratiques de déplacements automobilistes qu'elles prétendent pourtant réguler. Effectivement, tous les modèles précédents veulent adapter le



système à la demande sans cesse croissante et inégale (apparition des aires urbaines et étalement urbain). Cette croissance a engendré des problèmes de trafic et de capacité du réseau que ces modèles ne semblent pas vouloir réguler.

C'est également en 1975 que McFadden publie une théorie sur la demande en transport urbain. De son côté, il soulève le fait que les planificateurs en transport doivent se concentrer sur le TC qui est une solution aux problèmes de transport : *"transportation planners are focusing more and more attention on the development and improvement of public transit as an alternative to be explored and evaluated in alleviating various urban transportation problems. Selected improvements to the transit system might therefore divert enough persons from auto to transit to reduce highway traffic congestion and parking problems significantly. The need for urban highway construction, with its potential social disruptions might be reduced accordingly."* (Domencich & McFadden, 1975)

Les modèles antécédents ne sont pas suffisants pour l'analyse de la demande car ils ne sont pas adaptés aux nouveaux plans de transport. Le processus de décision de l'utilisateur comporte une infinité de variables telles que le temps, le coût, les conditions du déplacement mais aussi la préférence (par rapport à l'heure, à la météo, à l'humeur...) et la situation socio-économique de l'individu. La variable d'accessibilité au transport s'ajoute lorsque l'on traite le cas du TC. Néanmoins, le système de transport (sa politique) peut influencer sa décision. Un modèle de prévision de la demande en transport doit donc avoir une approche comportementale (des usagers) : *« which represents the decisions that consumers make when confronted with alternative choices »* (Domencich & McFadden, 1975) et doit être adapté au contexte de la planification (politique de transport). Il propose alors un modèle de la forme d'une fonction d'utilité additivement séparable dont les sous-fonctions sont les variables (indépendantes) influençant ou non le déplacement d'un individu donné puis il étend ce modèle à une population (agrégation) en intégrant des fonctions de probabilité pour des situations à choix binaire.

Avec l'augmentation des aires urbaines, l'étalement urbain, la congestion, la répartition complexe des entités sociales et économiques, la dégradation de l'environnement et bien d'autres problèmes que nous connaissons jusqu'à aujourd'hui sont apparus progressivement et ont pris plus ou moins d'ampleur. Manheim identifie trois dimensions critiques de changement pertinent au transport (Manheim, 1979) :

- le changement dans la demande de transport en termes de distribution spatiale et temporelle ;
- le changement dans les technologies disponibles ;
- le changement des valeurs pour les usagers et non usagers dans la prise de décision.

Kanafani développe une théorie à partir de l'approche basique de la demande en transport en milieu urbain (Kanafani, 1983). Il souligne que le besoin de se déplacer est directement lié et dérive des activités urbaines. Il est possible de stratifier les activités principales pour les ménages : travail, étude, magasinage ou loisir. Pour chacun de ces motifs, il est possible d'estimer la demande à partir des caractéristiques socio-économiques des ménages (taille, revenu, nombre d'actifs, etc.). Il explique que, selon les cas, toutes les activités ne peuvent être réalisées en raison de l'inexistence de lieux ou de leur inaccessibilité en termes de distance et de coût de transport. Le comportement d'un usager ne se limite donc pas au seul besoin d'effectuer une activité mais aussi au fait qu'il existe un moyen de transport qui lui permette de la réaliser. Ainsi, les caractéristiques socio-économiques permettent de définir la demande potentielle et les activités socio-économiques et le système de transport sont les déterminants du déplacement qui peuvent se montrer comme un obstacle au déplacement ou comme un choix supplémentaire.

Des guides de planification sont également publiés à partir des années 60. En 1984, Meyer et Miller publient un ouvrage qui explore les caractéristiques de la planification des transports urbains et qui développe un cadre de planification qui reflète le contexte actuel (requis et contraintes) des aires urbaines afin d'aider planificateurs et ingénieurs en

transport dans leurs prises de décision. Ils proposent un certain nombre de modèles interconnectés qui composent un système de modélisation du transport urbain (Meyer & Miller, 1984).

Nous faisons à nouveau un récapitulatif des avancées dans le domaine des transports dans le tableau suivant.

**Tableau 2.2: Récapitulatif des avancées en planification dans le domaine des transports urbains jusqu'en 1990**

Année	Chercheur	Avancée
1950	US HRB	Identification des Lignes de désir et paires OD
1952	Wardrop	Principe d'équilibre (affectation des déplacements)
1954	Mitchell, Rapkin	Méthodes empiriques sur les mouvements des personnes et des biens
1955	Beckmann, McGuire, Winsten	Paradigme micro-économique de l'offre et de la demande
1962	Oi, Shuldiner	Déterminants du déplacement (par motif)
1969	Wilson	Notion d'entropie dans la distribution spatiale des déplacements
1973	Hägerstrand	Notion des déplacements dans le contexte spatio-temporels
1975	McFadden	Théorie sur la demande en Transport Urbain
1979	Manheim	Dimensions critiques de changement pertinent au transport
1983	Kanafani	Mise à jour de la théorie de la demande en transport
1984	Meyer, Miller	Guide de planification pour le transport urbain

## 2.3 L'état de la recherche en planification

Comme nous l'avons soulevé précédemment, les planificateurs de transport sont à nouveau dans une période charnière où ils doivent modifier leurs objectifs pour résoudre les nouvelles problématiques : la congestion et les parkings dans les aires urbaines, l'environnement et la mentalité des usagers. Nous allons analyser dans cette dernière partie quelques articles récents où les chercheurs continuent de développer et d'adapter les modèles anciens au contexte actuel. C'est au début des années 90 que des nouvelles politiques de transport sont mises en place que ce soit au niveau local, régional ou fédéral.

Rappelons avant tout les travaux de Chapleau (Chapleau, 1986) qui propose une modélisation totalement désagrégée des déplacements. Cette approche s'appuie sur le concept d'itinéraire et de déplacement individuels. À partir d'enquêtes sur les déplacements, la base de données résultante est structurée de manière à pouvoir retracer les déplacements de chaque individu de chaque ménage enquêté. Le logiciel MADITUC (Modèle d'analyse désagrégée des itinéraires de transport urbain collectif) qu'il a développé est d'ailleurs le premier outil qui puisse retracer de tels déplacements sur le réseau TC (Chapleau et al., 1982). MADITUC et ses modules associés intègrent la modélisation graphique du réseau TC et de la demande (issue de la base de données de l'enquête), mais aussi la planification des ressources TC en fonction de l'offre de service désirée. Cependant, cette approche n'est pas applicable aux territoires où les données sont insuffisantes ou simplement inexistantes pour pouvoir modéliser des itinéraires aussi précis.

Dans le but d'interpréter l'utilisation de trois modes de déplacements par rapport aux caractéristiques de la population et à l'utilisation mixte du sol, Frank et Pivo (Frank & Pivo, 1994) ont effectué une étude sur les relations existantes (ou non) entre le choix modal et les facteurs de forme urbaine (comme la densité et l'aménagement du territoire) et non urbaine (caractéristiques de l'utilisateur, niveau de service TC etc.). À terme, cette analyse permettrait

de découvrir des moyens de planifier et manipuler des facteurs de forme urbaine qui puissent promouvoir l'accessibilité car une problématique importante de cette période était de diminuer les besoins de déplacements et la dépendance à l'automobile tout en augmentant la compétitivité entre les modes de déplacement.

Les formes urbaines sont les variables qui caractérisent la zone telles que la densité de population ou d'emplois, le nombre total de déplacements ou l'utilisation du sol. Les formes non urbaines sont les variables associées au niveau de service du système de transport ainsi que celles associées aux ménages et individus enquêtés : la tranche d'âge, la possession d'un permis de conduire, d'un abonnement de transport, d'un emploi.

Les données utilisées sont issues du Pudget Sound Transportation Project (PSTP) basé sur une enquête menée en 1989. Frank et Pivo ont émis des hypothèses à partir des descriptions statistiques fournies par le PSTP, par exemple l'existence d'une relation statistiquement significative entre le choix modal et les formes urbaines. Ils les ont vérifiées et ont identifié la nature des variables pertinentes par des analyses statistiques simples (corrélation, moyenne, régression). L'ensemble des résultats leur a permis de dresser un portrait du comportement de la population étudiée. On cite par exemple que les déplacements autos diminuent significativement lorsque la densité d'emplois est faible (entre 20 et 50 employés par acre).

Le second article que nous traitons concerne la seconde étape de la PSC : la distribution des déplacements (V. Arasan, 1996). Les auteurs ont pour objectif de définir les obstacles aux déplacements, en fonction de différents motifs et modes de transport, dans la ville de Tiruchirapalli en Inde. Réalisée à partir de l'enquête sur les ménages entreprise par le gouvernement, l'analyse des transports de cette ville a la particularité de considérer le mode « semi-public » (taxi, pousse-pousse) d'où une stratification en cinq modes de transport différents. Trois motifs ont été considérés : le travail, l'éducation et autres (loisirs, magasinage...).

Pour cette expérimentation, deux formulations de la méthode gravitaire (Reilly, 1931) ont été utilisées : l'une utilisant la fonction exponentielle, et l'autre a été élaborée par la division des routes publiques du département des transports des Etats Unis (Bureau of Public Roads, BPR). De plus, afin de déterminer les valeurs de coefficients les plus adéquats pour chaque formulation, il a fallu calibrer les modèles (à partir de l'optimisation des indicateurs de performance). Finalement, l'interprétation des variations de chaque modèle a été réalisée en prenant en compte la particularité du système étudié afin de donner une signification logique aux modèles. Pour que le modèle soit correct, les variations doivent être intuitives. Les indicateurs de performance sont également comparés pour la sélection de la meilleure formulation de modèle.

L'analyse de la distribution des déplacements dans la ville indienne a permis de mettre en évidence quelques effets de la distance entre le lieu de résidence et le lieu de destination d'un habitant :

- le frein au déplacement par rapport à la distance est plus fort pour le motif études que les autres motifs (les lieux d'études sont localisés à proximité de zones de résidence);
- le frein au déplacement par rapport à la distance est plus fort pour la marche à pied que pour le transport motorisé privé que pour le transport semi-public que pour le vélo que pour le transport public (la distance a peu d'effet sur l'usage du transport en commun contrairement à la marche à pied), ce qui souligne également l'effet du coût du transport au travers de l'usage moins spontané de la voiture que du transport en commun.

La critique que nous pourrions faire serait au niveau du choix des formulations des modèles. L'intégration d'un modèle de type américain paraît « inutile » devant le contraste existant entre les deux sociétés américaine et indienne ; cela s'observe dans les résultats

obtenus tout au long de l'expérience : le modèle BPR est toujours le moins satisfaisant. Il serait alors intéressant d'élaborer un processus d'analyse générale (et non stratifiée) de la distribution des déplacements propre au système indien, qui se baserait sur le comportement particulier de cette population. Ce modèle ne se baserait pas tout d'abord sur la distance (puisque'elle n'est pas pertinente pour tous les modes) mais sur le coût du transport. Puis, avant d'effectuer la calibration, il serait intéressant d'évaluer chaque indicateur statistique pour l'analyse stratifiée afin d'en dégager les indicateurs les plus significatifs. L'étape de calibration se multiplierait en plusieurs modèles afin de dégager un modèle optimal convergent. La validation du modèle s'appuierait également sur les résultats pour les données stratifiées afin d'évaluer la véracité pour chaque cas d'usage.

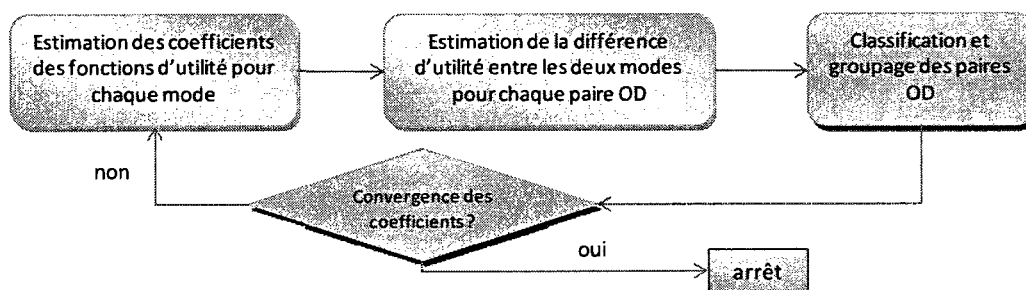
En 2004, Bonnel (Bonnel P. , 2004) s'inspire des modèles de choix modal développés par McFadden (1975) puis Ben-Akiva (1985) qu'il applique au cas de la ville de Lyon en France (les données empruntées sont celle de l'enquête à domicile effectuée en 1995-1996). Après l'étape de distribution des déplacements entre chaque zone dans la PSC, la répartition modale des déplacements est généralement effectuée à partir des modèles logit (basés sur la fonction logit). Sa méthode est utilisée ici lorsque seuls deux modes sont étudiés.

Même si la recherche se concentre aujourd'hui sur le développement des modèles désagrégés, qui interprèteraient plus fidèlement les comportements différents des usagers, les modèles agrégés restent plus d'usage notamment en France. En effet, malgré le haut degré d'incertitude des résultats et le niveau de précision insuffisant, la performance du modèle est plus grande.

Il propose alors une procédure itérative qui initialement se base sur des matrices OD les plus désagrégées possibles (issues de l'étape de distribution) et les coefficients de la fonction d'utilité pour chaque mode qui sont estimés à partir des variables explicatives du

système de zonage initial de l'étude. À partir de la classification des résultats pour chaque paire OD, il faut associer les paires où la différence de l'utilité entre les deux modes est très faible (« 50-50 » entre les deux modes). La procédure est alors réitérée pour le nouveau système de zonage jusqu'à ce que les coefficients convergent et que le seuil du nombre de déplacements à considérer (au minimum) atteint.

Ce regroupement des zones, qui suppose que l'individu aura la même propension pour un mode qu'il soit considéré dans sa seule zone ou parmi plusieurs zones, permet d'obtenir des résultats plus pertinents.



**Figure 2.2: Modifications apportées au processus de distribution par Bonnel, 2004**

Le diagramme résume la nouveauté qu'apporte Bonnel (en vert) au processus classique (en bleu). Il s'agit donc d'une agrégation d'un modèle désagrégé à la base afin d'en améliorer les performances. Malgré tout, les méthodes d'estimations sont toujours confrontées à deux problèmes contradictoires que les chercheurs tentent de résoudre : d'un côté le besoin de précision qui voudrait que le modèle soit le plus désagrégé possible pour pouvoir interpréter les résultats avec justesse et de l'autre la performance du modèle (exactitude) qui n'est obtenu que si l'on agrège les zones.

L'application s'est faite pour les déplacements pour motif travail sur une division de l'espace en 187 (plus petite échelle possible), 87 et 7 zones. Bonnel conclut de



l'expérimentation des méthodes classique et nouvelle sur les données de la ville de Lyon que l'agrégation est fortement recommandée lorsqu'on applique la méthode classique. D'ailleurs, celle-ci est suffisante lorsque le modèle est très fortement agrégé (7 zones). La méthode nouvelle a donné de meilleurs résultats lorsque l'échelle de zonage est plus petite (87 et 187 zones), ce qui permet d'obtenir un nombre suffisant d'observations modélisées.

La même année, le Centre de Recherche en Transport Urbain (Center of Urban Transportation Research) de l'Université de Floride Sud a été mandaté par le département des Transports de Floride (Florida Department of Transport) et Kittelson & Ass pour créer un modèle d'estimation de la demande agrégée (au niveau des arrêts) appliqué à Jacksonville et qui sera intégré au logiciel de prévision TLOS (Transit Level Of Service) créé par Kittelson & Ass.

L'article explique la méthodologie empruntée par le CRT dans la mise en place du contexte de l'étude au travers de l'exploitation du système de transport et des données de population dans le but d'identifier des variables pertinentes pour estimer la demande (Chu, 2004). Le modèle résultant doit décrire l'utilisation du TC pour un jour de semaine au niveau des arrêts avec 6 catégories de facteurs :

- les caractéristiques sociodémographiques dans la zone d'étude
- les valeurs requises par le logiciel TLOS
- l'environnement piétonnier (rue, croisement)
- l'accessibilité de la population et des emplois aux arrêts
- l'intégration du TC avec d'autres modes
- la compétition avec d'autres arrêts TLOS

Les estimations obtenues doivent pouvoir aider la société de transport à la modification des services existants ou à la proposition de nouveaux services. L'auteur recense et critique l'ensemble des alternatives de modélisation qui ont été abordées dans cette recherche. Il présente un certain nombre d'hypothèses et caractéristiques de la PSC qui montrent qu'il s'agit d'une méthode inefficace pour évaluer les impacts d'un changement du réseau ou du niveau de service ; par exemple, le découpage du territoire en zones (agrégation) et le manque de flexibilité au niveau des simulations (nécessité de reproduire toute la démarche de la PSC pour chaque situation). De plus, les nouveaux modèles qui s'appuient aujourd'hui sur les SIG (Systèmes d'Information Géographique) et l'accessibilité au TC présentent aussi des difficultés telles que la prise en compte des caractéristiques d'un arrêt, la représentation des effets compétitifs d'un réseau TC, la détermination simultanée nécessaire de l'achalandage et des niveaux de services.

L'analyse des problèmes liés à la collecte des données (activités aux arrêts, temps, caractéristiques démographiques, routes et direction) a permis de dresser un cadre d'identification des variables de prédiction potentielles.

Finalement, seules les recommandations à propos de la méthode ont été faites puisque le modèle choisi n'a pas été explicitement décrit. Cet article donne une base de réflexion qui permet aux planificateurs et aux chercheurs en transport d'aborder un certain nombre de choix à effectuer dans l'élaboration du processus d'estimation de la demande dans un système de transport particulier.

Sous la direction de Koppelman et Boyce, Bernardin a développé un modèle de prédiction innovant qui s'avère plus simple et moins coûteux que le processus classique (la PSC) et le processus basé sur les activités (« activity-based model »). Il publie un article présentant ce modèle basé sur l'accessibilité (« accessibility based model ») et le comparant aux anciens (V.L., 2008). L'auteur présente trois principales différences avec ses prédécesseurs :

1. le modèle est basé sur les variables d'accessibilité (il intègre les liens entre modes et déplacements);
2. il s'agit d'un modèle désagrégé et déterministe qui considère la variable de temps de manière plus continue;
3. pour les déplacements exogènes (non-home-based), le modèle considère l'origine et la destination.

Tandis que le modèle classique est fondé sur l'estimation du nombre de déplacements, celui-ci est basé sur les activités que l'individu doit ou veut réaliser. L'unité d'analyse n'est plus le déplacement mais plutôt le comportement de l'individu (emploi du temps, mode, lieu). Le but est de prédire quelles activités seront effectuées, à quel(s) endroit(s), pendant combien de temps, avec qui, quel mode de transport seront utilisés et, idéalement, le chemin emprunté lors des déplacements. Ce modèle se base sur la maximisation de l'utilité et sur des arbres de décisions qui traduisent le comportement des individus.

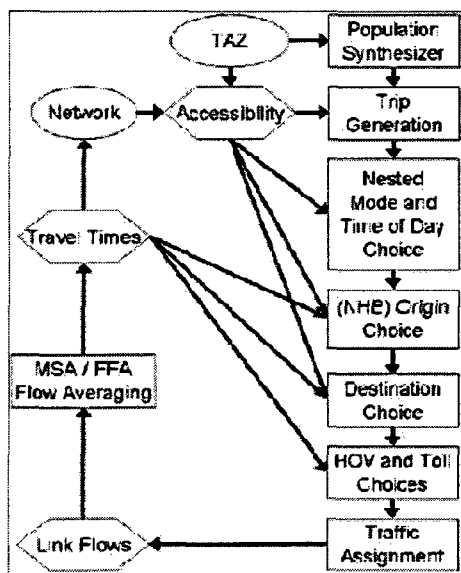


Figure 2.3: Modèle d'estimation de la demande de Bernardin (Bernardin, 2008)

Le modèle de Bernardin suit la procédure suivante :

1. La synthèse des données de population (collectées à partir du recensement et des enquêtes) afin qu'elles soient représentatives et interprétables.
2. La génération désagrégée des déplacements qui s'effectue à partir de données synthétiques de la population (données remaniées et représentatives de la population réelle issues du recensement et des enquêtes traditionnelles) et intègre les variables d'accessibilité; la régression linéaire est typiquement la méthode utilisée.
3. La répartition modale et temporelle (avant-distribution) des déplacements qui permet de garder le modèle désagrégé.
4. L'affectation des origines des déplacements « non-home-based » et l'affectation des destinations (distribution).
5. L'affectation des occupants des véhicules « High Occupancy Vehicle » et des péages; la méthode de maximisation de l'utilité est utilisée pour les étapes d'affectation.
6. L'affectation du trafic sur le réseau.

Malgré l'aspect performant du modèle théorique, l'article ne propose pas d'expérimentation effective à l'aide de données empiriques. L'auteur pèse simplement les avantages et désavantages de chaque étape de son processus en le comparant avec les modèles précédents. De plus, le processus demande une grande quantité de variables qui ne sont pas disponibles pour toutes les villes comme les données d'accessibilité et enquête sur l'occupation des véhicules et péages (sur lesquelles se joue l'innovation du modèle). La conclusion qui se détache de cet article est que ce modèle novateur se destine à des métropoles urbaines où le système de transport est déjà bien développé. Il permet d'attaquer des cas complexes de réseau où de nombreux paramètres doivent être pris en compte à la fois : multi-modes, complexité des routes, accessibilité.

En 2006, Peronne, Carr et Upton (consultants) publient un article dans lequel ils présentent une méthode alternative à la méthode classique d'estimation de la demande adaptée aux petites et moyennes villes (moins de 50000 habitants) afin d'effectuer une prévision des impacts du trafic sur le réseau (Perone, Carr, & Upton, 2006). Les modèles connus d'estimation de la demande sont utiles et efficaces mais leur intégration et leur entretien demandent un certain investissement que les petites communautés urbaines ne peuvent pas se permettre. De plus, les chercheurs n'ont toujours pas découvert de modèle qui soit directement transférable d'un réseau à un autre. C'est pourquoi, pour effectuer leur prévision, les petites et moyennes agglomérations se contentent de projeter les tendances historiques sur le futur.

Les auteurs proposent la méthode d'analyse cumulative en six étapes qui n'utilisent que des données disponibles dans ces agglomérations (tels que les données géographiques, les archives de recensement de la population et autres données historiques sur l'aménagement et l'utilisation du sol) :

1. Identification des prévisions de croissance par type d'aménagement (commerces, habitations, etc.) à partir de projections;
2. Affectation de la croissance attendue par analyse géographique et socio-économique du territoire;
3. Estimation des déplacements générés à partir de taux de génération (NCHRP Report 365: Travel Estimation Techniques for Urban Planning);
4. Détermination de la distribution des déplacements (modèle gravitaire);
5. Détermination des déplacements de transit (passage dans l'agglomération sans arrêt);
6. Affectation des déplacements (plus court chemin en tenant compte du facteur de congestion);

L'article présente simultanément une application pour la ville de Dallas en Orégon (12000 habitants) située à 15 milles de la capitale d'État et dont le quart des résidents effectuent au moins un déplacement vers la capitale par jour (notamment pour motif travail). La région, considérée comme le carrefour nord, est sujette à des problèmes de congestion car aucune route alternative n'est possible. L'application de la méthode d'analyse cumulative a permis de définir cinq zones de croissance à l'intérieur de la ville dans lesquelles de nouvelles constructions sont prédites (par type) dans un horizon de vingt ans. Cependant, les auteurs n'ont pas mentionné si l'étape d'affectation a mis en exergue les routes actuellement congestionnées dont la capacité devrait être augmentée ainsi que les zones d'intérêt par où pourraient passer des routes alternatives.

Finalement, l'article donne une méthode pratique et concrète pour les petites et moyennes villes. L'étude permettrait d'avoir un portrait de la ville à long terme afin que celle-ci puisse décider des investissements à favoriser dans le futur. Cependant, face au problème de congestion, les auteurs ne donnent pas de solution ou de conseil qui permettrait aux planificateurs de savoir à quelle étape et comment définir les alternatives pour les usagers en fonction de la politique du système de transport de la ville. Par ailleurs, le modèle ne différencie pas les modes de transports empruntés pour les déplacements générés et le transport de marchandises pour les déplacements de type transit.

Le dernier article pertinent pour cette partie est une comparaison entre les résultats modélisés et ceux observés dans l'étude de l'impact de l'intégration du métro de Copenhague sur le comportement des usagers du transport (G. Vuk, 2007). Les deux auteurs, Vuk et Jensen, ont voulu comparer les données observées disponibles à propos du comportement des usagers du métro à celles prédites par le modèle d'estimation de la demande des planificateurs de la ville.

Le métro de Copenhague est complètement automatisé, long de 16km et comporte 17 stations (dont 8 souterraines). La construction de celui-ci en 1994 et son ouverture en 2002 était attendue depuis plus de environ quarante ans. En effet, la ville (qui contient deux îles) connaissait des problèmes de congestion, d'environnement et de développement urbain.

Les données utilisées sont issues de données de comptages pour les véhicules privés et le TC ainsi que des enquêtes téléphoniques similaires aux enquêtes OD de Montréal. Le modèle d'estimation de la demande utilisé par la ville de Copenhague (divisée en 601 zones) prédit les déplacements pour un jour de travail moyen (pas le week-end) pour quatre motifs : travail, études, loisirs, navette. Il s'agit d'un modèle de choix discret appelé nested logit basé sur les étapes de la PSC. Les modes considérés sont : voiture, TC, vélo, marche. Pour le TC, sont considérés le bus, le train et le métro. L'affectation des déplacements s'appuie sur un modèle stochastique d'équilibre basé sur la fonction probit.

Les résultats concernant l'augmentation du trafic au bout d'un an (de 2002 à 2003) ont montré que le modèle a surestimé la croissance. Ceci est dû au fait que les facteurs géographiques et socio-économiques influençant les déplacements pour une telle ville n'ont pas tous pu être considérés et traduits par le modèle. De plus, cela montre que la période d'adaptation est plus longue que prévue par le modèle. La comparaison des données obtenues pour le choix du mode de transport sur deux ponts de la ville (point de passage critique) a donné des résultats similaires pour les comptages et pour le modèle : une baisse significative de l'utilisation de l'automobile et du bus au profit du métro.

Les auteurs concluent par des conseils dans l'application du processus d'estimation. Parfois, les zones à faible population et dont les profils sont très inégaux faussent les résultats dans l'étape de génération (« explosion » des prévisions). Il serait intéressant de réaliser la génération sur ces zones problématiques avant les autres pour réduire leur impact grâce à un facteur correctif. Un autre conseil serait d'inclure estimations désagrégées et prévisions désagrégées dans le modèle de génération. Enfin, l'utilisation du modèle basé

sur les activités (activity-based) serait pertinente car les usagers peuvent être sujets à emprunter le métro pour passer d'une activité à une autre.

## 2.4 Synthèse

Le tableau suivant récapitule les sujets récents de la recherche en transport qui ont été abordés. En conclusion, la compréhension et la modélisation du comportement de l'utilisateur par rapport à toutes les composantes qui caractérisent son déplacement – le moment, le motif, le mode, le transfert modal etc. – sont les challenges actuels des chercheurs qui permettraient de mesurer les réels impacts de la modification du système de transports.

**Tableau 2.3: Récapitulatif des avancées récentes en planification dans le domaine des transports urbains**

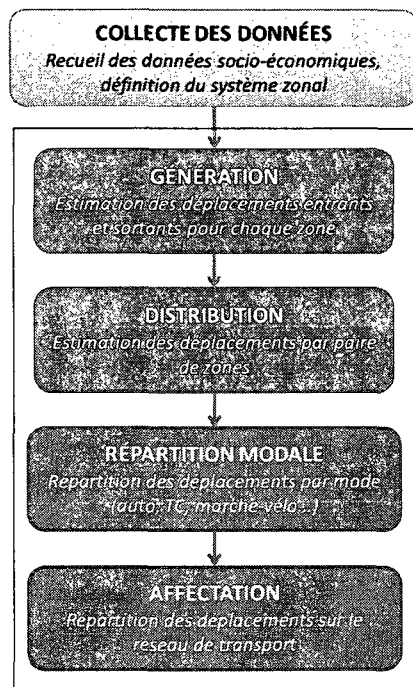
Année	Chercheur	Avancée
1994	Frank, Pivo	Relations entre le choix modal et les facteurs de forme urbaine (densité et l'aménagement du territoire) et non urbaine (caractéristiques de l'utilisateur, niveau de service TC etc.)
1996	Arasan	Les effets de la distance comme obstacle aux déplacements (en fonction de différents motifs et modes de transport)
2004	Bonnel	Procédure itérative pour la répartition modale des déplacements à partir de matrices OD
2004	Chu	Modèle d'estimation de la demande agrégée (au niveau des arrêts)
2006	Perone, Carr, Upton	Prévision des impacts du trafic sur le réseau pour les petites et moyennes villes (moins de 50000 habitants)
2007	Vuk, Jensen	Impacts de l'intégration du métro de Copenhague sur le comportement des usagers du transport
2008	Bernardin	Modèle d'estimation de la demande basé sur l'accessibilité



## **CHAPITRE 3. LA PROCÉDURE SÉQUENTIELLE CLASSIQUE**

Dans ce chapitre, nous nous appliquons à la présentation et à l'explication concise de la PSC. Pour chaque étape, il sera essentiel de faire un lien avec le transport en commun en indiquant les besoins du secteur, les données disponibles, les moyens généralement engagés, les résultats et leur application. Nous avons ajouté une étape à la PSC : la collecte de données, étape primordiale sans laquelle une modélisation de la demande ne peut être réalisée correctement. Les explications sont essentiellement basées sur l'ouvrage de Patrick Bonnel : « Prévision de la demande en transport » (2001). Nous concentrons nos efforts sur la génération et la distribution des déplacements. Les dernières étapes, qui sont la répartition modale et l'affectation, ne seront pas présentées avec autant de détails puisque dans le cadre de ce projet de recherche, nous ne considérons qu'un seul mode.

Le schéma ci-après récapitule les étapes de la procédure classique d'estimation de la demande en transport.



**Figure 3.1: Procédure Séquentielle Classique**

### 3.1 Collecte des données

Même si la collecte de données est rarement mentionnée dans la littérature, il s'agit d'une étape à part entière pour le processus de modélisation de la demande. Nous allons ici présenter les problèmes rencontrés lors de la collecte et l'adaptation des données.

Afin d'obtenir les variables potentiellement prédictives de la demande en transport, il faut tout d'abord définir les catégories de données pour les usagers mais aussi pour le système de transport :

- Socio démographie ou caractéristiques de la population (par ménage ou par personne): âge, sexe, occupation, revenu, possession d'une automobile.

- Aménagement du territoire : commerces, écoles, universités, zones industrielles, agricoles, d'habitation, centre d'affaires.
- Services de transport : réseau routier; pour le TC : nombre de lignes, nombre de passages, accessibilité (distance moyenne de l'arrêt, de la station de métro), niveau de service.
- Statistiques des usagers : déplacements types en pointe (comportement de mobilité).

Dans le Chapitre 4 de son ouvrage *Prévoir la Demande en Transport*, Bonnel présente une méthodologie de production de données soit pour analyse seule ou alors pour modélisation et menant à la mise en place d'une enquête.

Le principal problème associé à la production et la collecte de données est celui de la représentativité dans l'espace et dans le temps. Sur un segment de route, les infrastructures et les usagers n'auront pas tous le même profil. De même, dans une journée ou dans une semaine, les usagers feront des déplacements différents (se rendre au travail, retourner au domicile, aller à l'école, aller faire des courses...). Il est donc important, lors de la préparation d'une enquête, de cibler la population (lorsque seule une partie est enquêtée) et de bien définir le cadre conceptuel c'est-à-dire des éléments influant sur l'étude. Bonnel rappelle que « dans le passé, les études de prévision de trafic avaient souvent pour finalité première le dimensionnement de nouvelles infrastructures de transport, il était donc nécessaires de produire une matrice OD correspondant à la période de pointe de la demande de transport. Pour le TC, l'intensité maximale de la pointe du matin (...) était assez souvent privilégiée. L'évolution des préoccupations de planification et d'évaluation socio-économiques nécessite de plus en plus de modélisation concernant l'ensemble de la journée (GESMAD, 2000), voire des périodes plus longues prenant en compte les déplacements de fin de semaine. » (Bonnel, 2001)

Ainsi, plus les données sont fines et variées, plus la modélisation représentera précisément les usagers mais plus celle-ci sera complexe devant la grande quantité de données et la multitude de variables. Généralement, les données sont agrégées afin de faciliter la modélisation. Le comportement moyen est estimé et doit être interprété en fonction de l'erreur et du niveau d'agrégation.

L'application de cette étape au TC se voit simplifiée car on ne considère qu'un mode et qu'aujourd'hui la plupart des régions urbaines possèdent des données statistiques sur le TC. Même si les enquêtes OD et/ou à bord sont réalisées dans les grandes villes européennes (Paris, Londres, Stockholm...) et américaines (New York, Montreal, Mexico...), le choix des données TC à intégrer au modèle reste fonction des autres types de données comme nous l'explique l'article du FDOT.

Par ailleurs, malgré la pertinence d'une variable explicative, elle peut être omise car elle ne correspond pas au contexte et à l'objectif du modèle. Par exemple, dans notre cas d'étude du comportement de l'usager du TC, il ne sera pas nécessaire d'intégrer des variables qui concernent le type d'habitation (année de construction, taille, bâtiment) ou bien les lignes TC spécifiques empruntées pour chaque déplacement (le chemin serait alors propre à chaque usager, un niveau de détails trop important ici).

### **3.2 Génération**

L'étape de génération de la demande de déplacements consiste à déterminer le nombre de déplacements réalisés par les résidents du secteur d'étude. Pour chaque zone du secteur, il s'agit d'estimer le nombre de déplacements entrant et sortant. Sont utilisés les termes d'émissions ou de productions (déplacements dont l'origine se situe dans la zone) et d'attractions (déplacements dont la destination se situe dans la zone). Parfois, les déplacements sont générés par motif (on parle de segmentation par motif) pour améliorer la

significativité du modèle. Les motifs généralement considérés sont : le travail, les études, le magasinage, les loisirs et les autres motifs.

Rappelons que nous nous intéressons aux déplacements endogènes, c'est-à-dire aux déplacements des usagers résidant dans la zone d'étude. Nous ne considérons pas les déplacements attirés par l'extérieur du terrain d'étude ou émis vers l'extérieur.

Les principales familles de modèle de générations sont :

- Les modèles normatifs (utilisant des normes de génération faute de données);
- Les modèles par régression linéaire;
- Les modèles par classification croisée.

### **3.2.1 Le système de zonage : Agrégation-Désagrégation**

Comme nous l'avons mentionné plus haut, le problème de division du territoire est critique pour le modèle car c'est notamment en fonction du zonage que la performance du modèle sera élevée ou non et que l'interprétation des résultats sera plus ou moins difficile. En fonction du niveau de détail requis, il est nécessaire de trouver un compromis pour que l'agrégation soit minimale afin d'obtenir un pourcentage d'erreur le plus acceptable possible pour que l'interprétation des résultats reste pertinente.

### **3.2.2 Méthode la plus utilisée : la Régression Linéaire**

La méthode la plus simple et la plus courante est celle de la régression linéaire multiple. On considère que le trafic est généré par l'utilisation du sol. En d'autres termes, les caractéristiques liées à l'utilisation du sol peuvent expliquer les variations du trafic. Il est alors possible d'estimer le nombre de déplacements à partir des valeurs caractéristiques d'une zone donnée (Douglas & Lewis, 1971).

Nous ne rappellerons pas le concept de régression dans ce mémoire mais nous présentons simplement le modèle résultant :

$$Y = \sum_{j=1}^n \beta_j X_j + \varepsilon \quad (1)$$

où :

$n$  : nombre de variables indépendantes utilisées

$Y$  : taux de déplacements (nombre de déplacements par personne) issu de données de référence

$\beta_j$  : coefficients de la régression à déterminer

$X_j$  : variables indépendantes (issues du recensement, des données réseau)

$\varepsilon$  : erreur, variable négligeable proche de 0

Il est possible d'obtenir un modèle qui estime le nombre de déplacements pour une zone entière, pour un ménage ou pour un individu. Le nombre de déplacements pour une zone est insuffisant pour interpréter la demande puisque l'aménagement du sol n'est pas homogène (même pour un segment de route). À l'origine, la modélisation s'effectuait sur les ménages puisque la mobilité des individus y est interdépendante (partage des moyens). Aujourd'hui, des facteurs, tels que l'augmentation du nombre de ménages à une personne ou la diversification des profils de ménages qui complexifie la modélisation, conduisent les planificateurs à considérer l'individu directement comme unité de mesure.

Bonnell cite les variables explicatives qui se retrouvent le plus couramment dans les modèles de génération. « Pour l'émission, les variables qui sont prises en compte sont le plus souvent relatives aux caractéristiques de la population résidant dans la zone considérée. Les variables que l'on retrouve le plus souvent dans le modèle de génération sont les suivantes :

- revenu du ménage (parfois de l'individu);

- motorisation du ménage généralement exprimé en nombre de voitures disponibles ou à la disposition du ménage (...) par individu du ménage ou par individu de plus de 18 ans du ménage;
- composition du ménage lorsque la génération s'effectue au niveau du ménage;
- nombre de personnes ou de ménages de la zone;
- nombre d'enfants en âge d'aller à l'école en distinguant souvent selon l'âge afin de retrouver les niveaux de scolarité;
- l'âge des individus, ou plus généralement une répartition par classe d'âge.

Pour l'attraction, on utilise davantage les variables qui caractérisent les activités disponibles dans la zone (y compris la population comme source d'attraction par exemple pour les loisirs (visites à la famille, aux amis...)) :

- nombre d'emplois de la zone, souvent segmentés en emplois tertiaires et non tertiaires et parfois en isolant les emplois commerciaux parmi les emplois tertiaires;
- nombre de places scolaires;
- nombre de ménages ou d'individus de la zone. » (Bonnell, 2001)

Ces variables sont celles que nous intégrerons en priorité dans notre modèle d'estimation.

### **3.2.3 Régression par classe (ou par catégorie)**

Lorsque la diversité des comportements est trop grande, il est intéressant de définir des classes de ménage ou d'individu qui permettent d'obtenir des modèles qui expliquent plus précisément le comportement de l'utilisateur. Avant d'effectuer l'étape de génération par régression linéaire, il faut déterminer les classes d'individus (ou de ménages) aux comportements homogènes par une étude préliminaire des données. Cependant, il est nécessaire que la taille de l'échantillon étudié soit suffisante pour effectuer la régression par

classe et, sachant qu'il s'agira d'un modèle de prévision, il est important d'intégrer des variables traduisant l'évolution de la classe. L'intégration de variables muettes est préconisée (variable catégorique indiquant l'appartenance à la classe).

### 3.2.4 Les indicateurs de performance

L'indicateur principal est le coefficient de détermination  $R^2$  qui représente le pourcentage de variance expliquée par le modèle de régression.  $R^2$  est défini comme le rapport entre la moyenne des carrés des écarts à la droite de régression sur la moyenne des écarts à la valeur moyenne de la variable expliquée généralement appelée  $Y$ .

Le test F (de Fisher) sur le modèle complet permet de tester l'hypothèse suivante : « tous les coefficients de la régression à déterminer sont égaux et nuls ». Il s'agit du rapport entre les carrés expliqués et les carrés non expliqués, pondéré par les degrés de liberté et les résidus. Plus la valeur de F est grande, plus le pouvoir explicatif du modèle est grand.

### 3.2.5 Critique

L'étape de génération telle que nous l'avons présentée paraît simple à réaliser mais les limites de la régression montrent certaines lacunes auxquelles il est nécessaire de remédier. Cette partie nous permet de donner des recommandations quant à la préparation de la modélisation.

Tout d'abord, les hypothèses du modèle de régression peuvent impliquer l'adaptation des variables :

- La régression linéaire suppose que chaque variable « exerce une influence linéaire sur le phénomène », ce qui est rarement le cas en réalité. L'étude de la colinéarité est donc importante avant l'intégration ou non d'une variable dans le modèle de régression.



- L'homoscédasticité c'est-à-dire la constance de la variance du nombre de déplacements prédit (l'erreur est toujours la même). Il est donc important de garder une certaine homogénéité dans le découpage du terrain d'étude et de préconiser les taux (ratio) aux nombres totaux.
- La multicollinéarité entre les variables explicatives (les variables sont corrélées entre elles) qui peut fausser l'interprétation des résultats et diminuer la performance du modèle. L'étude préliminaire des corrélations pour l'ensemble de variables est primordiale.

Les estimations résultantes représentent les marges de la matrice OD qui sera remplie à l'étape de distribution des déplacements. Pour que celle-ci soit cohérente, il est indispensable que la somme de toutes les émissions soit égale à la somme de toutes les attractions. Cette situation est difficile à obtenir à partir du modèle de régression seul. Ainsi, des facteurs de pondérations sont créés pour les déplacements attirés (on considère que les déplacements émis sont plus fiables) tels que :

$$A_{jpondéré} = A_j \times \frac{\sum_i E_i}{\sum_j A_j} \quad (2)$$

et donc :

$$\sum_i E_i = \sum_j A_{jpondéré} \quad (3)$$

Enfin, le problème de la stabilité des paramètres est mis en évidence lorsque le planificateur veut effectuer une prévision. Les coefficients sont estimés pour une période donnée et pour un terrain d'étude particulier, l'évolution socio-économique temporelle peut les faire varier. La seule solution est d'effectuer une projection ou d'émettre des hypothèses d'évolution (par défaut de données) pour les coefficients.

### 3.3 Distribution

La seconde étape de la PSC consiste à construire la matrice Origine-Destination des déplacements d'une population. Après avoir généré la demande (émissions  $E_i$  et attractions  $A_j$ ) pour chaque zone  $i$ , il est nécessaire de répartir les déplacements par paire de zones  $(i,j)$  en fonction des secteurs accessibles et en fonction de leur attractivité et du profil de la population.

Ci-après, nous avons modélisé une matrice OD avec les contraintes aux marges générées par l'étape de génération des déplacements :

		Zones de destination				Emissions
		1	2	...j...	n	
Zones d'origine	1	$T_{11}$	$T_{12}$	$T_{1j}$	$T_{1n}$	$E_1 = \sum_{j=1}^n T_{1j}$
	2	$T_{21}$	$T_{22}$	$T_{2j}$	$T_{2n}$	$E_2 = \sum_{j=1}^n T_{2j}$
	...i...	$T_{i1}$	$T_{i2}$	$T_{ij}$	$T_{in}$	$E_i = \sum_{j=1}^n T_{ij}$
	n	$T_{n1}$	$T_{n2}$	$T_{nj}$	$T_{nn}$	$E_n = \sum_{j=1}^n T_{nj}$
Attractions		$A_1 = \sum_{i=1}^n T_{i1}$	$A_2 = \sum_{i=1}^n T_{i2}$	$A_j = \sum_{i=1}^n T_{ij}$	$A_n = \sum_{i=1}^n T_{in}$	$T = \sum_{i=1}^n A_i = \sum_{i=1}^n E_i$

Figure 3.2: Matrice Origine-Destination avec les contraintes aux marges

L'étape de distribution consiste à estimer le nombre de déplacements ( $T_{ij}$ ) qui s'effectuent entre deux secteurs donnés. Il doit pouvoir intégrer les différentes dimensions qui définissent et motivent un déplacement telles que les données :

- d'aménagement du territoire (points d'attraction, accessibilité au réseau),
- du trafic (zones de trafic),
- de population (caractéristiques des ménages, âges, occupations...),
- du service (pour le TC : nombre de lignes, d'arrêts...).

Trois types de technique analytiques sont présentés ici :

- celle basée sur les facteurs de croissance,
- celle dite « synthétique » qui utilise les lois physiques (modèles gravitaire),
- et enfin, le modèle d'opportunité basé sur les probabilités.

### **3.3.1 Le modèle à facteurs de croissance**

La technique consiste à estimer les déplacements à partir d'une matrice de référence (historique). La matrice de référence est transformée afin de satisfaire les contraintes aux marges. Les transformations peuvent être effectuées :

- soit à partir d'un facteur uniforme, on multiplie la matrice de référence par une constante estimée pour obtenir la matrice OD finale,
- soit à partir de facteurs de croissance, pour chaque ligne et chaque colonne est estimé un facteur à partir du modèle de référence qui est alors appliqué à chaque valeur de la ligne ou de la colonne (peut se faire par itération pour les lignes puis les colonnes avec la méthode de Furness (Furness, 1965) ou par itération en simultanée avec la méthode de Fratar (Salter, 1974)).

Ce type de modèle est simple en termes de calculs mais demande une matrice de référence, matrice que l'on ne peut pas toujours trouver et qui n'est pas forcément de bonne qualité, ce qui risque d'affecter les performances du modèle final. Par exemple, le manque de données pour remplir la matrice de référence (absence d'informations pour certaines paires de zones) et qui serait représenté par des cases nulles (aucun déplacement effectif entre les deux zones) fausserait l'estimation. De plus, l'utilisation d'une matrice de référence sous-entend que les nouveaux déplacements qui sont répartis utilisent les moyens (infrastructures, aménagements) de la matrice de référence. En ce sens, il est plus difficile de tenir compte de la présence d'un nouveau service ou d'une nouvelle infrastructure.

### 3.3.2 Le modèle gravitaire

La modélisation la plus connue, qui s'inspire de la loi de Newton, utilise une fonction de résistance (l'impédance) au déplacement qui est souvent définie à partir de la distance : plus le secteur est éloigné du point d'habitation, moins l'utilisateur sera enclin à s'y rendre. On doit effectivement considérer un paramètre influent dans la motivation d'un déplacement : le coût du déplacement. Celui-ci peut être interprété : en terme de distance, de temps, d'argent ou même d'une quantité (de population par exemple). Le modèle gravitaire a pour forme générale :

$$T_{ij} = \alpha_i E_i \beta_j A_j f_{ij} \quad (4)$$

où :

$T_{ij}$  est le nombre de déplacements entre la zone  $i$  et la zone  $j$

$\alpha_i$  et  $\beta_j$  sont des coefficients interdépendants mis à jour à chaque itération qui traduisent le respect des contraintes aux marges tels que  $\alpha_i = \frac{1}{\sum_{j=1}^n \beta_j A_j f_{ij}}$  (5) et que

$$\beta_j = \frac{1}{\sum_{i=1}^n \alpha_i E_i f_{ij}} \quad (6)$$

$E_i$  le nombre total de déplacements générés par la zone  $i$

$A_j$  le nombre total de déplacements attirés par la zone  $j$

$f_{ij}$  le coût de déplacement considéré (généralement une fonction de la distance  $d_{ij}$  ou du coût généralisé moyen  $C_{gij}$ ) entre les deux zones

La fonction d'impédance est notée  $f_{ij}$ . Il existe une infinité de formulations possibles de ce modèle selon le choix du coût considéré et donc de la fonction d'impédance dont les trois plus connues (ici appliquées à la distance) sont :

- la fonction puissance :  $f_{ij} = d_{ij}^{-k}$  (7);
- la fonction exponentielle :  $f_{ij} = \exp(-k \cdot d_{ij})$  (8);
- la fonction combinée exponentielle puissance :  $f_{ij} = d_{ij}^{-a} \cdot \exp(-b \cdot d_{ij})$  (9).

Le ou les coefficients sont à déterminer et leur estimation constitue le calage du modèle. Le calage du modèle permet d'intégrer des coefficients spéciaux pour des paires OD dont le comportement des usagers est particulier (attraction ou génération particulière des déplacements).

Le coût généralisé est une fonction qui représente le coût perçu par l'utilisateur et que celui-ci attribue à son transport, son temps de parcours et d'accessibilité au transport. Ainsi, la prise de décision concernant le mode et le chemin emprunté pour le déplacement correspondra au coût généralisé minimal. Les valeurs de chaque composante de ce coût sont des valeurs moyennes qui correspondent à des usagers moyens. La fonction est alors définie comme suit (version la plus simple) :  $C_g = P + T \cdot h$  (10) où  $P$  est le prix du transport,  $T$  est le temps de transport (temps de parcours et temps d'accès) et  $h$  est la valeur du temps pour l'utilisateur. La composante du temps peut également être divisée en deux pour différencier l'accessibilité au réseau et le parcours en lui-même.

### 3.3.3 Le modèle d'opportunité

Le modèle d'opportunité « considère que les individus choisissent la destination la plus proche de leur lieu d'origine, leur offrant l'opportunité de satisfaire leur besoin (de déplacement ou d'activité). (...) L'application de la méthode consiste donc à énumérer les

destinations possibles (en fait l'ensemble des zones) et à les classer par ordre croissant de distance. La probabilité qu'un individu choisisse une zone est donc égale à la probabilité qu'il ne choisisse pas les zones plus proches multipliée par la probabilité de trouver des opportunités susceptibles de le satisfaire dans la zone. » (Bonnell, 2001)

Le modèle se base donc sur la mise en place d'un ensemble de probabilités de satisfaction et de non satisfaction où l'estimation des distances doit être suffisamment précise pour assurer la qualité des résultats. Dans le cas du transport en commun, la mesure des distances de parcours est un point critique car elle doit se baser sur les lignes de TC (trajet, correspondances, transfert modal) en plus du réseau routier de base.

### 3.3.4 Les indicateurs de performance

Il est important d'apprécier chaque formulation du modèle gravitaire à partir d'indicateurs de performance. Deux types de mesures statistiques permettent d'évaluer un modèle :

- la qualité de l'ajustement cellule par cellule (« goodness-of-fit ») ;
- la comparaison macroscopique avec la distance/le coût/le temps de parcours réel(le).

Trois indicateurs de type « goodness-of-fit » ont été utilisés ici : le  $\phi$  normalisé (divergence), l'erreur moyenne absolue normalisée et le coefficient  $R^2$ .

### 3.3.5 Critique

La première critique que nous pourrions émettre est la nécessité d'une matrice de référence dans les deux méthodologies pour pouvoir ajuster les résultats. L'inexistence ou la mauvaise qualité de cette matrice engendrerait des valeurs estimées erronées. Il serait intéressant de trouver une technique qui ne se base pas sur une matrice de référence mais sur des caractéristiques socio-économiques.

De plus, la fonction d'impédance doit être définie correctement pour ne pas fausser les valeurs des matrices OD. Lorsque l'on ne dispose pas du coût généralisé, faute de données, la distance est une variable pertinente qui doit être estimée intelligemment. La plupart des programmes de calcul estiment les distances et temps des plus courts chemins à vitesse moyenne, ce qui correspond à un parcours idéal hors pointe (au mieux). Si les données associées aux déplacements correspondent à la pointe du matin, les résultats peuvent être mal estimés dans la matrice de distribution.

Enfin, contrairement à l'étape de génération, la distribution des déplacements est complexe car elle nécessite de nombreuses estimations en plus du nombre de déplacements et des itérations. Le programme utilisé pour compléter cette étape doit pouvoir traiter une matrice de très grande taille (au moins 100 zones pour les modèles agrégés soit plus de 10000 valeurs à estimer sans compter les coefficients et contraintes).

### **3.4 Répartition Modale**

Cette quatrième étape utilise les matrices OD issues de la distribution des déplacements. Il s'agit ici de répartir les déplacements en fonction du choix du mode de déplacement de l'utilisateur parmi la séquence des choix possibles. A l'issue de cette étape, nous obtenons des matrices OD par mode.

La présente recherche s'intéressant uniquement au mode TC, nous n'aborderons pas en détails cette étape de la PSC. Il est simplement important de retenir que la répartition modale se base généralement sur la régression logistique qui implique le coût généralisé (défini à partir du prix du transport, des temps de parcours et d'accès) dans sa fonction d'utilité. Le coût généralisé est la formalisation de la fonction de l'offre (Manheim, 1979).

### 3.5 Affectation

La dernière étape de la PSC consiste à affecter les déplacements sur le réseau à partir des matrices OD. « Le principe de base de tous les modèles d'affectation est de considérer que l'individu cherche à minimiser le coût généralisé de son déplacement pour une origine-destination et un mode donnés. » (Bonnell, 2001) Un modèle d'affectation peut prendre en compte les deux dimensions essentielles : la contrainte de capacité (des routes) et l'utilité déterministe (tous les individus perçoivent le coût d'un itinéraire de manière identique) ou probabiliste (intégration des facteurs du choix qui ne sont pas dans le coût généralisé).

Différentes méthodes sont utilisées pour chaque combinaison de dimensions (avec ou sans contrainte de capacité, type d'utilité). La méthode déterministe la plus simple est l'affectation tout-ou-rien et se base sur le fait que les individus empruntent forcément le plus court chemin pour aller d'une origine à une destination. L'utilisation d'algorithmes de plus courts chemins (Dijkstra, 1959) est de rigueur.

### 3.6 Outils de planification

Un grand nombre d'outils de modélisation existent sur le marché. Nous allons ici présenter trois logiciels courants qui utilisent des méthodes différentes pour estimer la demande.

#### 3.6.1 GNE7/QRSII (AJH Associates)

L'ensemble de logiciels est un outil de planification puissant basé sur la PSC dont le paramétrage est appliqué aux villes américaines. Il permet de lancer les quatre étapes de la procédure d'estimation de la demande sur un système de transport préalablement défini (géographie, données socio-économiques, répartition modale).

General Network Editor (GNE) est le Système d'Information Géographique qui permet la codification du réseau de transport (plusieurs modes peuvent être inclus). Il permet de



définir le découpage territorial désiré en zones d'analyse (TAZ : Travel Analysis Zones), avec un maximum de 15 zones dans la version gratuite. Chacune de ces zones est définie par un centroïde, caractérisé par différents paramètres socio-économiques clefs comme le nombre de ménages ou le revenu moyen. Ces centroïdes sont reliés aux nœuds du réseau principal par des connecteurs de centroïdes, qui représentent le temps d'accès à ce dernier. Le logiciel permet alors de relier les différentes TAZ entre elles. Les liens obtenus, qui peuvent représenter le réseau routier et/ou celui de transport en commun, sont aussi définis par différents paramètres tels que la capacité ou la vitesse moyenne des véhicules l'empruntant.

Quick Response System (QRS) exécute la PSC sur le réseau généré par GNE. Le programme demande également des caractéristiques socio-économiques globales qui lui permettent d'estimer les coefficients du modèle. Les hypothèses de base du programme ne prennent en compte que le nombre de ménages et d'emplois par zone et les comportements moyens des utilisateurs du réseau entier (comme le nombre moyen de déplacements par ménage, la répartition modale et par motif de déplacement) sans différencier les taux pour chaque zone d'étude.

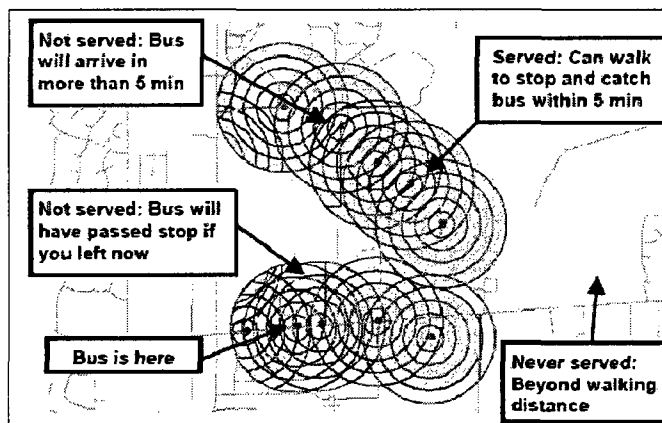
Les informations entrées ajustent les coefficients par rapport au système étudié. Si les données fournies sont incomplètes, le logiciel intégrera des valeurs par défaut (valeurs types pour une ville américaine). Il est également possible de modifier manuellement ces coefficients en arrière plan du programme.

### **3.6.2 TLOS**

Le logiciel Transit Level Of Service (TLOS) a été développé par le Département des Transports de Floride en conséquence du projet d'amélioration des données analytiques des systèmes de TC en Floride. L'issue de ce projet a été la création d'un indicateur de mesure de l'accessibilité au TC intégrant :

- la couverture du réseau;
- les fréquences de services;
- les heures de service;
- les chemins empruntés par les piétons jusqu'aux arrêts et leur qualité;
- la densité de population et d'emploi.

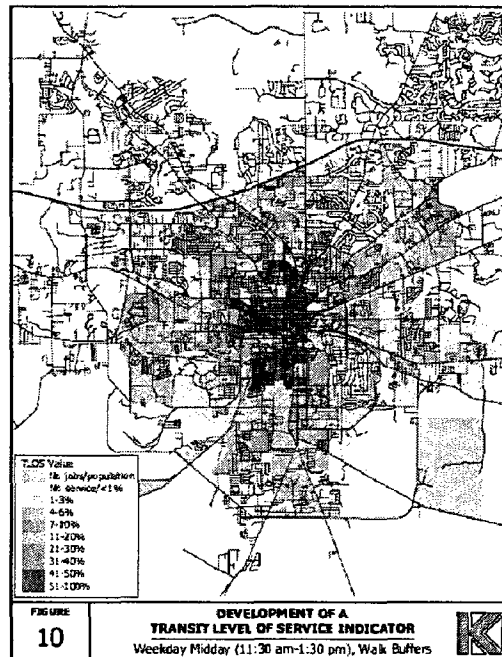
Ainsi, à l'origine, le concept de base du logiciel est de déterminer pour une zone, une heure et un bus donnés, l'accessibilité à cette ligne. Sur le schéma ci-après un exemple d'application est expliqué. Chaque cercle représente un temps de marche d'une minute par rapport au cercle inclus (le plus large) dans celui-ci et chaque point représente un arrêt.



**Figure 3.3: Exemple d'application simple du logiciel TLOS issu du guide de l'utilisateur (Kittelson & Associates, Inc., 2001)**

Pour estimer la demande à l'intérieur d'un buffer (temps de parcours d'une minute à pied jusqu'à l'arrêt), le logiciel se base simplement sur le nombre de résidents et d'emplois de la

zone en considérant qu'ils sont uniformément répartis sur la zone (TAZ) et qu'il sont tous usagers du TC. Ainsi, l'indicateur TLOS est calculé en fonction du temps de parcours de l'utilisateur jusqu'à l'arrêt mais aussi par rapport aux usagers.



**Figure 3.4: Exemple de résultats obtenus par TLOS pour une région urbaine en pointe du midi (Kittelsson & Associates, Inc., 2001)**

Ce logiciel permet aux planificateurs d'évaluer le niveau de service de leur système de transport pour une période donnée mais aussi de simuler des scénarios de modification ou de mise en place d'un nouveau service ainsi que de changements au niveau sociodémographique (augmentation de la population par exemple). TLOS peut également faire un état des accès aux arrêts de TC (qualité des routes). De manière indirecte, il peut aider à l'étude de la congestion, expliquer la répartition modale de secteurs donnés et de fait aider à améliorer le système de transport.

### 3.6.3 TBEST

Le Transit Boarding Estimation and Simulation Tool (TBEST) a également été développé par le FDOT. Il s'agit de la troisième génération d'outils de prévision de la demande TC. Le modèle de la demande à l'arrêt estime le nombre de déplacements (embarquement direct et transfert) par route, par direction et par heure. Le logiciel se base sur le nombre de tournées, sur les caractéristiques des buffers aux arrêts, l'accessibilité et les effets d'une configuration alternative de la ligne ou du réseau. Il utilise les données du recensement de la population pour caractériser un arrêt : le pourcentage de ménages ne possédant pas de véhicule personnel, le pourcentage d'afro-américains dans la population, le revenu moyen du ménage, les données d'emploi. Il considère également les données TC telles que la fréquence de service, l'accessibilité et la connectivité du réseau. TBEST peut aussi être utilisé pour effectuer une modélisation agrégée de la demande à l'échelle des routes ou des zones à partir des résultats désagrégés.

Le choix d'une telle désagrégation n'a pas été dû au besoin de précision mais au fait que la source d'accès prédominante au TC est la marche. En conséquence, les variables d'emploi et de population ainsi que l'accessibilité aux arrêts sont des facteurs importants pour expliquer et caractériser la marche et l'utilisation du TC.

Tout comme QRS, des coefficients standards (créés à partir des données de la ville de Jacksonville en Floride en 2000) sont utilisés et peuvent être modifiés par l'utilisateur. Il est possible d'imposer un scénario (modification du service) et le logiciel pourra en mesurer les effets. Par ailleurs, il peut effectuer des analyses plus poussées pour des routes ou des secteurs précis et, à partir de données réelles (de comptage), il peut ajuster les résultats préalablement obtenus pour valider le modèle de base (avant de lancer les prévisions).

### 3.7 Synthèse

La PSC est la méthode la plus répandue pour modéliser la demande en transport. Ce chapitre a permis de présenter les concepts théoriques de base associés à chaque étape du processus ainsi que quelques outils concrets appliquant plus ou moins le processus. Ces logiciels montrent bien l'importance des données pour la qualité de la prévision : même si la procédure d'estimation est correcte, la pertinence des variables, leur qualité et leur quantité est primordiale pour que le modèle soit performant. La suite du mémoire se consacre au travail de recherche pour le développement d'une méthode hybride qui puisse s'adapter par rapport

## CHAPITRE 4. MÉTHODOLOGIE

L'objectif de cette recherche est de proposer un modèle d'estimation de la demande qui permette de réaliser des matrices OD décrivant des déplacements TC pour un découpage donné. Le remplissage de cette matrice s'effectue à partir de données démographiques, socio-économiques et du réseau TC de base pour qu'il puisse s'appliquer à des territoires différents.

Le problème d'agrégation est majeur dans le sens où le planificateur souhaite connaître le comportement des usagers avec le plus de précision possible pour pouvoir mieux appréhender les déplacements. Dans le modèle classique, plus un modèle de génération est agrégé plus les résultats obtenus sont proches de la réalité. Cela est dû au fait que les comportements particuliers des usagers ne sont plus visibles puisque les variations sont moins nombreuses (il y a moins de zones). En ce sens, plus les zones sont petites (et plus elles sont nombreuses), plus les déplacements particuliers sont considérés et complexifient les estimations du modèle. Cependant, le besoin de détails nous impose une certaine agrégation de même que le niveau d'erreur du modèle et la limite des outils (pour la distribution des déplacements notamment) imposent une révision du processus classique. La méthodologie empruntée doit pouvoir s'adapter aux moyens du projet tout en gardant un niveau de détails correct dans les résultats. Nous avons donc défini une méthode hybride qui se base sur la désagrégation progressive du modèle.

Ce chapitre concerne la méthodologie que nous proposons pour la génération et la distribution des déplacements mais aussi le traitement préliminaire des données. Après une présentation rapide des outils du projet, nous abordons les concepts de la méthode hybride par rapport à la méthode classique. Chaque étape de la procédure sera exposée avec plus de précisions et les résultats de l'expérimentation seront présentés dans les chapitres suivants. Le diagramme ci-après décrit le processus général de la méthode hybride combinant les deux étapes modifiées de la PSC : génération et distribution des déplacements. La collecte des données détermine la complexité du problème. En fonction des variables recueillies et

du système zonal, il faut agréger plus ou moins le modèle et le temps de résolution en devient plus ou moins long.

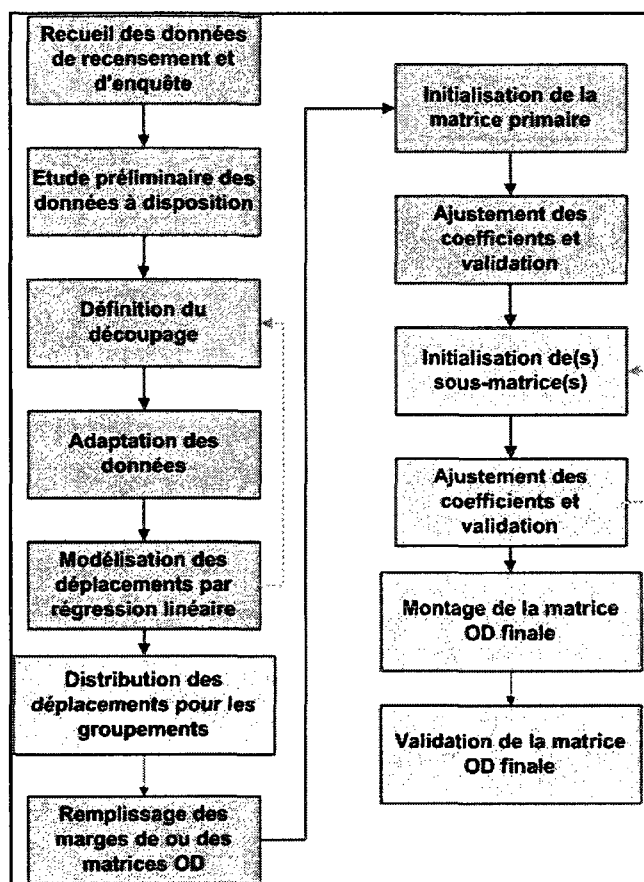


Figure 4.1: Méthode hybride d'estimation de la demande TC

## 4.1 Outils du projet

### 4.1.1 MICROSOFT OFFICE EXCEL

Le tableur de la suite Office présente toutes les qualités de base recherchées pour réaliser le projet : grande capacité de stockage de données, outils statistiques (régression), outils

d'analyse de simulation (solveur pour problème sous contraintes). Il peut donc être utilisé pour toutes les étapes de modélisation, notamment dans une perspective de démonstration et prototypage d'un modèle.

Cependant, les limites (en quantité de données) de cet outil nous forcent à agréger le modèle. Le solveur peut estimer un maximum de 200 variables.

#### **4.1.2 FGIS (© Université 2003 du Wisconsin)**

À l'origine, Forestry Geographic Information System (FGIS) a été conçu pour des études forestières mais s'applique à tout type de surface et en tant que Système d'Information Géographique permet de modéliser un territoire et ses divisions, d'y calculer des distances et d'y représenter des variables caractéristiques telles que des données de population ou des points d'intérêts.

Cet outil nous permet de faire l'analyse préliminaire des données et d'interpréter les résultats de modélisation obtenus au travers de la représentation des déplacements des usagers du TC.

#### **4.1.3 STATA (STATACORP LP)**

Ce logiciel d'analyse statistique permet d'effectuer toutes les procédures statistiques de notre étude. Par conversion des données, le programme permet d'obtenir des résultats complets directement interprétables pour la régression et complète parfaitement l'outil Excel.

#### **4.1.4 MICROSOFT MAPPOINT**

Ce logiciel a essentiellement été utilisé pour son outil de calcul de distance et temps de parcours. Le programme estime, à partir des coordonnées latitude-longitude de l'origine et de la destination, le chemin optimal parcouru en voiture. Ce chemin est donc de distance et de temps les plus courts et prend en compte la vitesse moyenne des routes empruntées.



## **4.2 Collecte et traitement des données**

### **4.2.1 Définition du système de zonage : agrégation-désagrégation**

Le système zonal doit être le plus désagrégé possible pour que le modèle soit le plus descriptif possible. Un modèle plus fin permet une représentation plus significative du comportement des usagers pour une meilleure compréhension de leurs déplacements. Il sera de ce fait possible d'obtenir une modélisation intégrée des déplacements pour tous les systèmes zonaux.

Même si le système zonal se base originalement sur la sectorisation utilisée pour les données de recensement de la population, les enquêtes OD ou celles utilisées par la système de transport existant, le choix du découpage doit se faire en fonction du besoin de détail du planificateur. Il s'agira alors de regrouper les zones « similaires » pour pouvoir modéliser correctement.

Le regroupement des zones doit s'effectuer en fonction de l'espace étudié, du contexte d'étude et des résultats de la méthode classique à différentes échelles de zonage.

Plusieurs critères peuvent aider à choisir les zones à regrouper :

- les infrastructures (industries, logements, commerce, pas d'exploitation) ;
- l'importance du réseau TC (densité, accessibilité) ;
- le profil particulier des ménages (donc des usagers du TC) par rapport aux autres zones ;
- le profil particulier d'emplois (zones du centre-ville (CV)).

Ainsi, lorsque plusieurs zones présentent une similitude flagrante (zone industrielle ou commerciale ou résidentielle, CV), le regroupement de celles-ci est préconisé.

Dans notre expérimentation, nous agrégerons uniquement les zones du CV et ce pour les raisons suivantes :

- le CV est composé de plusieurs zones de recensement ;
- les déplacements se destinant au CV et hors-CV ont globalement des profils différents, notamment dans le cas des métropoles, il serait donc pertinent de regrouper les zones centrales ;
- le réseau TC de ces zones est particulièrement dense et est en décalage avec celui des zones hors-CV, il est important de considérer ces zones plus en détails.

La définition des zones et de l'agrégation est l'étape primordiale pour la méthode hybride car elle détermine la performance du modèle. Le choix du découpage se fait en fonction de la disponibilité des données. Dans le cas de Montréal, l'agrégation a été suggérée par le système zonal du recensement. Cependant, même si le regroupement des zones permet d'avoir statistiquement de meilleurs résultats, ceux-ci ne sont pas intéressants pour l'interprétation car trop globaux. Dans un premier temps, nous avons utilisé cette méthode pour le Centre Ville. En effet, le CV est une zone fondamentale car il s'agit du centre névralgique des activités donc des destinations des déplacements, notamment pour le motif travail. De plus, le comportement des usagers se destinant hors-CV diffère grandement de celui des usagers se destinant au CV.

#### **4.2.2 Analyse préliminaire des variables**

Avant de réaliser la génération, il est nécessaire de regrouper et réorganiser préalablement les données de manière à avoir, pour chaque zone du découpage, une donnée de chaque

variable. Il est d'abord intéressant d'étudier les corrélations existantes entre les variables c'est-à-dire de savoir si elles auraient des variations globalement similaires (ce qui veut dire qu'une variable donnée puisse être définie par une autre par une fonction linéaire simple «  $Y=aX$  ») ou non. Les matrices de corrélations permettent d'écarter dès le départ des variables trop linéairement reliées pour la régression. En effet, la présence de telles variables dans un modèle suppose que les coefficients estimés sont erronés à cause de l'influence « similaire » de ces variables sur la variable à modéliser.

L'étude statistique mais aussi géographique est pertinente. L'analyse de la répartition des données au niveau géographique aide à l'agrégation du territoire de manière uniforme par rapport aux variables explicatives.

Les indicateurs de mobilité sont les données associées aux déplacements, les variables à expliquer : le nombre de déplacements entrant et sortant par sexe, par motif et par âge, et les caractéristiques du réseau TC. Leur répartition comparée à celle des données de population permet de définir un système de zones plus ou moins pertinent.

#### **4.2.3 Stratification des variables**

L'analyse préliminaire consiste ensuite à modifier les variables pour améliorer leur significativité pour le modèle. Pour chaque variable explicative, il s'agit d'analyser les tendances globales par rapport à la variable de déplacements (à expliquer) en émettant les hypothèses suivantes :

- lorsque la tendance est similaire (parallélisme) et linéaire, la variable est pertinente et explique une partie des variations des déplacements;
- lorsque la tendance est similaire et non linéaire (variations), la création de plusieurs classes pour la variable pourrait améliorer le modèle;

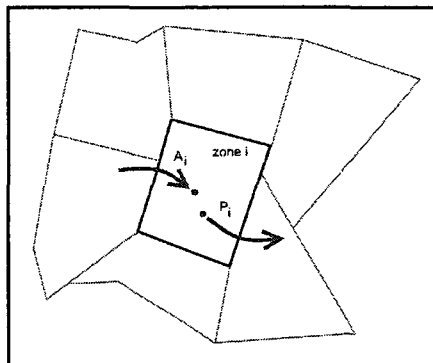
- lorsqu'une partie des données présentent une tendance similaire, la création d'une nouvelle variable est préconisée;
- lorsque la tendance n'est pas similaire, la variable n'est pas pertinente.

L'étude des variables indépendantes qui seront intégrées au modèle de génération est importante pour en augmenter les performances puisqu'il s'agit des variables qui seront potentiellement utilisées pour expliquer les déplacements. L'observation graphique permet donc d'identifier les classes associées aux variables. Au final, sont obtenues de nouvelles variables qui modélisent les déplacements avec plus de précision.

### **4.3 Génération des déplacements**

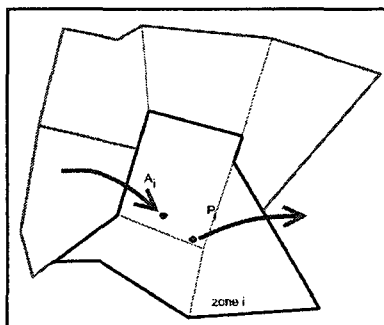
La première étape de la Procédure Séquentielle Classique consiste à estimer les déplacements produits et attirés pour chaque zone d'étude. Cette étape de génération est essentielle puisqu'elle définit le nombre global de déplacements par motif (travail, études, loisirs, magasinage).

La méthode classique définit un modèle de régression linéaire multiple qui estime le nombre de déplacements attirés  $A_i$  ou émis  $P_i$  par une zone  $i$  à partir des caractéristiques de la population, de l'aménagement et du réseau de transport de celle-ci. Cependant, devant la difficulté de modéliser le comportement des usagers du TC à un certain degré d'agrégation, les résultats ne sont pas toujours satisfaisants et se traduisent par des modèles statistiquement peu explicatifs et/ou pas assez pertinents (les variables ne sont pas significatives).

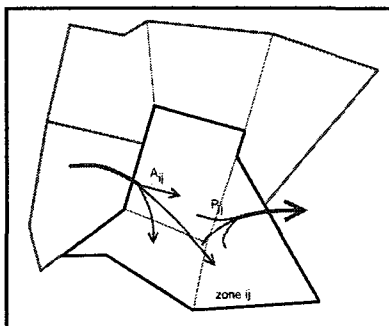


**Figure 4.2: Première étape de la PSC - Génération de la demande**

Les données de recensement sont disponibles pour différents types de zonage. Il est donc possible d'appliquer la méthode classique à des échelles d'étude variées. Cependant, il est évident que plus la désagrégation n'est forte, plus le modèle devient complexe et plus ses performances diminuent. La méthode hybride doit donc consister à réaliser un modèle de régression pour des ensembles de zones à partir duquel les déplacements générés sont distribués par zone. L'agrégation préliminaire permet d'avoir des estimations plus globales et plus pertinentes (proches de la réalité) afin d'améliorer l'affinement des résultats de la génération des déplacements par zone. Les données concernant la population sont considérées en taux (quantité par personne et non par zone).



**Figure 4.3: Première étape de génération des déplacements**



**Figure 4.4: Seconde étape de génération des déplacements**

A la différence de la méthode classique qui propose des coefficients « uniques » pour tous les types de zones, nous pouvons à présent différencier des espaces où le comportement de la population est particulier tel que le centre ville ou un quartier résidentiel.

Dans le logiciel QRSII, la calibration des coefficients se fait à partir de données statistiques pour l'ensemble des zones (part modale, nombre moyen de voyages par personne). Il ne s'ajuste pas par rapport à des données d'enquête sur les déplacements effectifs de la population.

On peut donc supposer que les résultats seront meilleurs que pour la méthode classique puisqu'on impose plus de précision au niveau des coefficients : ils sont créés pour un nombre restreint de zones présentant des profils de population similaires. Leur estimation est simplifiée et la performance du modèle est plus grande. Bien sûr, la performance de l'estimation des déplacements pour chaque zone dépend aussi de la pertinence du résultat obtenu pour le groupement.

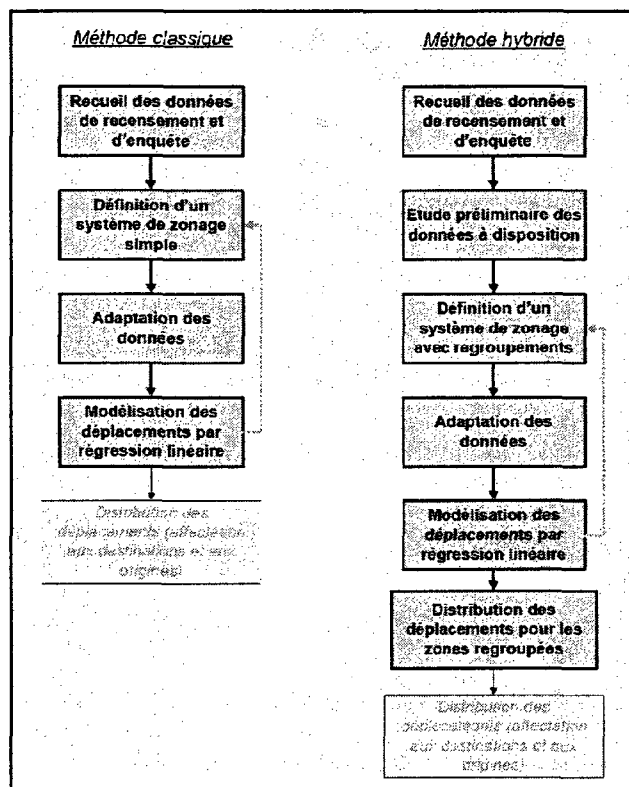
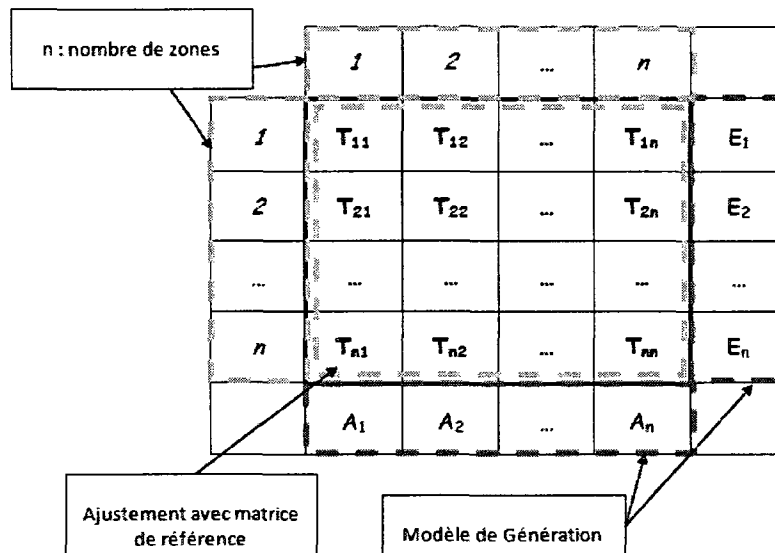


Figure 4.5: Comparaison des méthodes de génération classique et hybride

## 4.4 Distribution des déplacements

Le problème associé à l'étape de distribution est essentiellement lié à la limite de l'outil par rapport à la taille de la matrice. Nous allons donc, tout comme l'étape précédente, échelonner la distribution en divisant la matrice en plusieurs sous-matrices que l'on remplit au fur et à mesure.

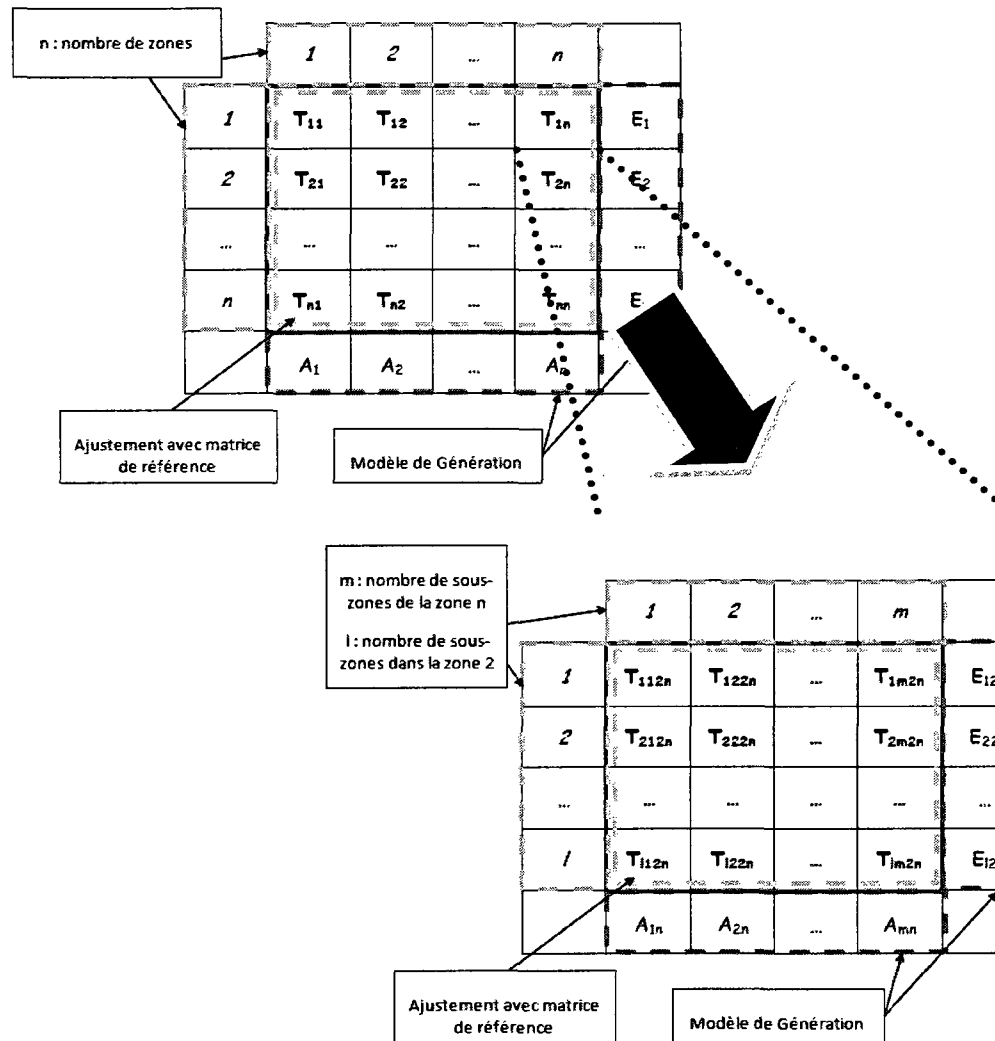
Auparavant, la taille de la matrice était fixée par rapport au découpage initial. Les contraintes aux marges étaient déterminées par l'étape de génération. Les déplacements estimés dans la matrice étaient ajustés en fonction d'une matrice de référence.



**Figure 4.6: Matrice de distribution des déplacements**

À présent, puisque nous disposons de contraintes aux marges pour des matrices OD de tailles différentes, il est possible d'affiner les données de distribution par étape. La première étape sera de remplir la matrice OD où le système de zonage est le plus agrégé. Pour chaque case, nous obtiendrons le nombre total de déplacements à redistribuer qui constitue une contrainte supplémentaire pour la distribution et ainsi de suite.





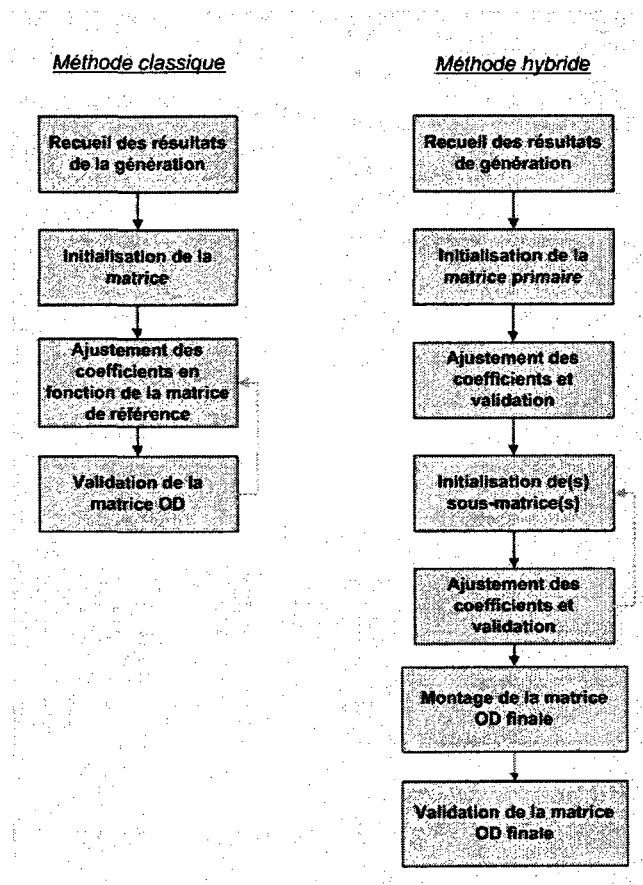
**Figure 4.7: Distribution des déplacements pour la méthode hybride**

Cet affinage par étape des résultats permet d'estimer plus précisément les déplacements et de pouvoir interpréter les résultats de différents points de vue (à grande comme à petite échelle). La localisation plus précise des déplacements permet au planificateur d'évaluer

plus rapidement le réseau et de prendre des décisions plus claires. Bien sûr, les résultats doivent être validés à partir de la matrice de référence et de l'analyse avec les données socio-économiques que nous avons suggérée dans le chapitre précédent.

La méthode hybride est plus longue mais plus efficace que la méthode classique puisqu'elle permet d'obtenir des résultats détaillés. Elle ajoute un grand nombre de problèmes à résoudre comparativement à l'unique problème de la PSC.

Les résultats obtenus ne seront pas du même ordre car le modèle hybride va plus loin. Ils ne seront donc pas vraiment comparables puisque la distribution classique correspond à la première étape de ce nouveau modèle.



**Figure 4.8: Comparaison des méthodes classique et hybride de distribution des déplacements**

## 4.5 Synthèse

La méthode hybride est une alternative à la PSC puisqu'elle intègre différentes échelles de résultats, ce qui permet de réaliser différents types d'interprétations en fonction du besoin de l'utilisateur.

La collecte de données est l'étape primordiale car la pertinence des résultats dépendra de la qualité des chiffres recueillis. L'analyse préliminaire et la stratification des variables aident à adapter les variables existantes au contexte de l'estimation de la demande en TC.

L'étape de génération modélise les déplacements entrants et sortants de chaque zone et sous-zones d'étude par régression linéaire. Les coefficients des modèles (par découpage, par sexe, par type de destination) sont définis à partir des variables précédemment créées.

La distribution génère les matrices OD par étape. La méthode classique demande effectivement l'utilisation d'un solveur à grande capacité et, même en limitant le nombre de zones, les résultats de la distribution ne sont pas pertinents. Ainsi, l'application par échelle de découpage permet de définir progressivement les déplacements par paire OD avec moins d'erreur et une résolution plus rapide.

Par la suite, chacune de ces trois étapes sera détaillée et illustrée par quelques résultats de l'étude de cas pour l'île de Montréal.

## **CHAPITRE 5. TRAITEMENT DES DONNÉES**

Trois types de données sont collectés pour l'analyse préliminaire qui permettra de les unifier pour la modélisation : les données du recensement réalisé en 2000, de l'enquête OD de 2003 et des caractéristiques du réseau TC en pointe du matin. Il s'agit ensuite de transformer les variables pour pouvoir croiser les données. Nous avons également calculé, à partir des coordonnées longitude-latitude du centroïde de chaque zone, la distance au CV (croisement des rues Peel et Sainte-Catherine) à partir du logiciel Mappoint.

### **5.1 Sources de données**

#### **5.1.1 Territoire**

Vu la multitude de sources de données relatives à la modélisation de la demande, il est important de se doter de données spatialisées sur les différents découpages territoriaux, ne serait-ce que pour fusionner et rendre les données compatibles entre elles. Pour toutes les sources, les données recueillies suivent des découpages administratifs dont les coordonnées sont disponibles. Les systèmes d'information géographiques (SIG) sont tout indiqués pour accueillir les données spatialisées. Nous avons pu recueillir les données géographiques du territoire d'étude grâce à l'initiative de démocratisation des données (IDD) du site de la CRÉPUQ (Conférence des Recteurs et des Principaux des Universités du Québec).

#### **5.1.2 Recensement**

Les données de recensement contiennent différentes informations sur la population pour un territoire donné. Des variables démographiques et socio-économiques sont disponibles telles que : le nombre d'habitants, l'âge et le sexe, la taille et la composition des ménages, le type de logement, le statut (familiale, emploi), le revenu, la possession automobile, la langue et l'origine ethnique.

Les découpages territoriaux du recensement canadiens sont relativement fins, allant jusqu'à des unités d'environ 1000 personnes; des microdonnées sont également disponibles. À des

fins de compatibilité, le découpage zonal utilisé pour la modélisation respecte habituellement le découpage de l'organisme de statistiques, qui lui-même s'assure de respecter les découpages administratifs (municipalités, régions).

Les données collectées seront les variables explicatives des déplacements c'est-à-dire qu'elles serviront à estimer le nombre de déplacements générés dans la première étape de la PSC.

Les données de recensement que nous avons utilisées pour la recherche sont disponibles sur le site de statistiques du Canada (<http://www.statcan.gc.ca>). Le site propose différentes unités géographiques normalisées dont les deux pertinentes pour notre étude :

- les subdivisions ou secteurs de recensement (plus de 850 lieux) ;
- les divisions de recensement (moins de 30 lieux).

Il est évident que les divisions de recensement ne sont pas pertinentes pour l'étude, c'est pourquoi nous choisissons de travailler avec les subdivisions que nous appelons SR. La ville de Montréal compte exactement 930 secteurs pour ce recensement de l'année 2001 dont 543 sur l'île de Montréal. Pour chaque secteur nous avons recueilli les informations brutes suivantes :

- superficie et densité du secteur ;
- population totale, par sexe, tranche d'âge, statut, origine, langue parlée, occupation, moyen de transport, revenu, niveau d'études, religion ;
- nombre de ménages par type, taille, revenu ;
- nombre de logements par type, taille, ancienneté ;
- taux d'activité, d'emploi, de chômage.

### 5.1.3 Caractéristiques du réseau

Ces données permettent de caractériser l'offre de transport collectif dans un secteur donné. Pour le système zonal de l'étude, il s'agit de connaître le nombre d'arrêts, le nombre moyen de passages de véhicules à ces arrêts, les modes faisant partie de l'offre de service, etc. Ces données servent à mieux estimer le nombre de déplacements de transport collectif induits.

Les caractéristiques du réseau sont fournies par l'opérateur. Les caractéristiques du réseau peuvent être directement obtenues à partir du plan du réseau tandis que les données sur les passagers sont obtenues à partir d'enquêtes, de comptages ou de données statistiques des planificateurs du TC. La disponibilité de ces données sur les passagers peut donc varier d'une région à une autre. Dans la grande région de Montréal, l'Agence métropolitaine de transport centralise ces données dans le cadre, entre autres, de la tenue de l'enquête régionale origine-destination.

Nous avons pu recueillir des données concernant le service de TC sur l'île de Montréal telles que :

- le nombre de lignes desservant le SR durant la période de pointe du matin ;
- le nombre d'arrêts localisés à l'intérieur du SR et desservis durant la période de pointe du matin ;
- le nombre de voyages traversant le SR durant la période de pointe du matin ;
- le nombre de voyages de métro qui traversent le secteur durant la période de pointe du matin ;
- la distance à la station de métro.

Ces données ont été collectées pour l'heure de pointe du matin car c'est une période de fort achalandage et donc les déplacements sont majoritairement endogènes, ce qui est important pour notre étude.

#### 5.1.4 Enquêtes origine-destination

Les enquêtes ménages OD sont effectuées par téléphone ou en personne auprès de la population. Les données récoltées touchent le ménage (taille, possession automobile, lieu de domicile), les personnes (âge, sexe, statut, permis de conduire) et les déplacements (heure, origine, destination, modes, motifs, utilisation des ponts, stationnement, etc.).

Par téléphone, un intervieweur équipé d'un logiciel spécialisé (CATI, *computer assisted telephone interview*) collecte de l'information sur les déplacements effectués par un ménage lors de la journée précédente. Les enquêtes OD à domicile, qui consistent également à recueillir les données de déplacement de la veille, peuvent s'effectuer en même temps que le recensement ou seules. Dans la grande région de Montréal, depuis 1970, il existe une tradition quinquennale de tenue de ces enquêtes par téléphone, à un taux d'échantillonnage d'environ 5 à 7% (Chapleau et Lavigueur, 1991).

En ce qui concerne les données issues de l'enquête Origine-Destination effectuée en 2003 que nous utilisons dans nos expérimentations, nous avons uniquement recueilli le nombre de déplacements entre un SR Origine et un SR Destination donnés par SR de résidence, motif, sexe et classe d'âge. Ces données servent essentiellement pour la matrice de référence du modèle.

Rappelons qu'il existe six catégories de motif : le travail (1), les études(2), les loisirs (3), le magasinage (4), autres motifs (5), le retour (6) ; et que les classes d'âge sont définies par tranche de 5 ans (cohorte) : 0-4ans (1), 5-9ans (2), 10-14ans (3) etc.



## **5.2 Structures de données proposées**

### **5.2.1 Données territoriales**

Pour les villes du Canada, le site de la CRÉPUQ propose des produits géographiques du recensement au Canada. Les fichiers sur les limites spatiales pour différents niveaux de système zonal (subdivision, secteur, aire de recensement) sont proposés pour toutes les régions métropolitaines et agglomérations de recensement au Canada en format ArcInfo et MapInfo pour les trois derniers recensements (2006, 2001, 1996). Il s'agit de cartes géographiques avec les limites de chaque zone et les coordonnées géographiques (longitude, latitude) de leurs centroïdes. Elles permettent alors de calculer grossièrement les distances entre les centroïdes.

### **5.2.2 Données opérationnelles**

Des données de recensement qui se retrouvent généralement sur les sites internet gouvernementaux de statistiques et des données du réseau TC ont été utilisées. Elles sont collectées pour des échelles de zones identiques.

### **5.2.3 Données sur les paires origine-destination**

Les données les plus pertinentes et exploitables sont les résultats d'enquête Origine Destination. Les informations nécessaires pour la modélisation des déplacements que nous avons collectées sont les suivantes :

- les données géographiques du domicile de l'utilisateur, de l'origine et de la destination du déplacement (les références des zones correspondantes);
- le sexe et l'âge (ou la cohorte) de la personne;
- le motif du déplacement : travail, études, loisirs, magasinage, autre ou retour.

Ces données servent de base à la création du modèle qui doit générer les déplacements entrant et sortant d'un lieu donné de la région d'étude.

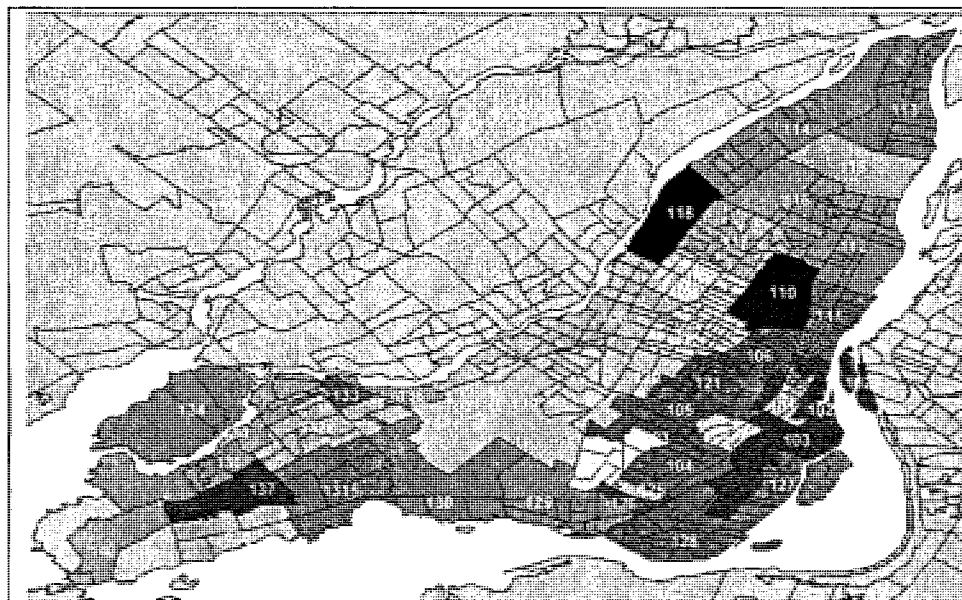
### **5.3 Choix du système zonal**

En ce qui concerne la ville de Montréal, le secteur de recensement SR est le système zonal le plus désagrégué dont nous disposons soit plus de 800 zones. Nous disposons également des SM ou subdivisions municipales avec une centaine de zones.

Le système de zonage primaire sera du type secteur municipal (SM). Au nombre de 40 sur l'île, les résultats de la méthode classique pour l'étape de génération seront performants mais leur interprétation ne permettra pas une vision suffisamment détaillée du réseau.

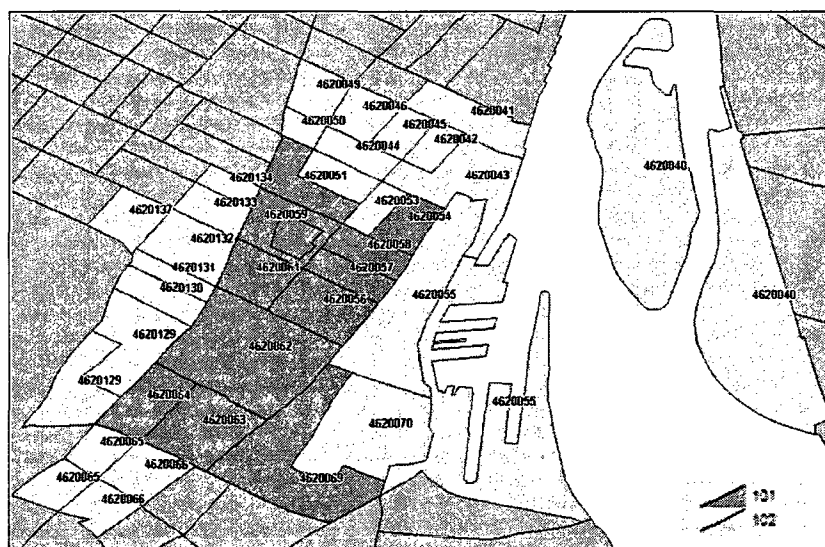
Le CV ne comporte que deux zones, ce qui n'est pas suffisant pour l'interprétation des résultats pour le mode TC puisque ce territoire représente le centre névralgique du réseau de transport public à Montréal et nécessite le plus de détails possibles au niveau du comportement des usagers.

Nous avons donc choisi d'utiliser le regroupement des zones de type secteurs de recensement qui représentent alors un secteur municipal (ou plutôt deux) pour la méthode hybride. Le découpage obtenu a été modélisé géographiquement sur la figure 5.1. Les couleurs délimitent les SM et les lignes délimitent les SR.



**Figure 5.1: Découpage territoriale en secteurs municipaux de l'Île de Montréal**

Par la suite, la méthode hybride va diviser les SM représentant le CV en 38 SR.



**Figure 5.2: Découpage secondaire du Centre Ville de Montréal (SM101 et SM102)**

## **5.4 Analyse préliminaire : classification des variables**

L'analyse de régression telle que nous l'avons expliquée correspond à la génération des déplacements c'est-à-dire le remplissage des marges des matrices OD du modèle décrivant les déplacements entrant et sortant d'une zone d'étude. Il est nécessaire de regrouper et réorganiser préalablement les données de manière à avoir pour chaque zone du découpage une donnée de chaque variable. L'étape de génération peut s'effectuer directement sur Excel (dans la limite de 200 variables) ou sur un logiciel de statistiques qui permettrait alors d'obtenir plus de détails au niveau analytique. Regroupées en un tableau d'extension DTA (pour STATA) ou STA (pour STATISTICA), les données sont compilées et les modèles de régression sont prêts à être construits.

Avant de réaliser les modèles, il est intéressant d'étudier les corrélations existantes entre les variables c'est-à-dire de savoir si elles présentent des variations globalement similaires ou non. Les matrices de corrélations permettent d'écarter dès le départ des variables trop linéairement reliées pour la régression.

### **5.4.1 Premières Transformations des variables**

Il est nécessaire d'effectuer des transformations au niveau des données puisque celles-ci ne sont pas exploitables telles quelles, notamment du fait que les enquêtes OD donnent le nombre de déplacements par triplet de SR (d'origine, de destination et de résidence) tandis que les données de recensement concernent chaque SR individuellement.

Pour chaque SR d'origine, nous avons par exemple calculé :

- la totalité des déplacements endogènes par profil d'utilisateur (âge, sexe, motif) ;
- le nombre moyen de déplacements par habitant du SR (ratio totalité des déplacements sur nombre d'habitants du SR) ;

- les déplacements ayant pour motif le travail et les études ;
- les déplacements émis et attirés par la totalité des résidents du SR, par sexe, par motif.

Nous totalisons alors une centaine de variables et 930 cas (nombre de SR) sur lesquelles nous allons effectuer différentes analyses de régression. Par ailleurs, nous ne retiendrons que les données de l'île de Montréal, nous pouvons donc omettre les cas des SR des autres régions. Il ne reste donc plus que 543 cas.

Le tableau ci-après décrit chaque variable dont nous disposons.

Tableau 5.1: Variables de déplacements à modéliser

Variable	Description
srorig	Référence du SR d'origine
sexe	{1, 2, total}
hors_cv_i_j	taux de déplacement vers hors CV pour motif = i et COHORT = j
hors_cv_i	taux de déplacement vers hors CV pour motif = i (toutes COHORT)
total_hors	taux de déplacement vers hors CV (tous motifs, toutes COHORT)
total_hors-6	taux de déplacement vers hors CV (tous motifs sauf 6=retour, toutes COHORT)
vers_cv_i_j	taux de déplacement vers CV pour motif = i et COHORT = j
vers_cv_i	taux de déplacement vers CV pour motif = i (toutes COHORT)
vers_cv_i-6	taux de déplacement vers CV pour motif = i (toutes COHORT)
total_verse	taux de déplacement vers CV (tous motifs, toutes COHORT)
total_verse-6	taux de déplacement vers CV (tous motifs sauf 6=retour, toutes COHORT)
total	taux de déplacement total (tous motifs, toutes COHORT)
total_mimo	taux de déplacement (tous motifs sauf 6=retour, toutes COHORT)

Pour un motif i (de 1 à 5) et une tranche d'âge donnée j (de 1 à 18), le taux de déplacement « hors\_cv\_i\_j » (« vers\_cv\_i\_j ») représente le nombre moyen de déplacements effectués par un habitant du secteur de la tranche d'âge donnée vers un secteur hors du centre ville (vers un secteur du centre ville) pour le motif i.

Pour un motif  $i$ , le taux «  $\text{hors\_cv\_t}\sim i$  » («  $\text{vers\_cv\_t}\sim i$  » ) représente le nombre moyen de déplacements effectués par un habitant du secteur vers un secteur hors du centre ville (vers un secteur du centre ville) pour le motif  $i$

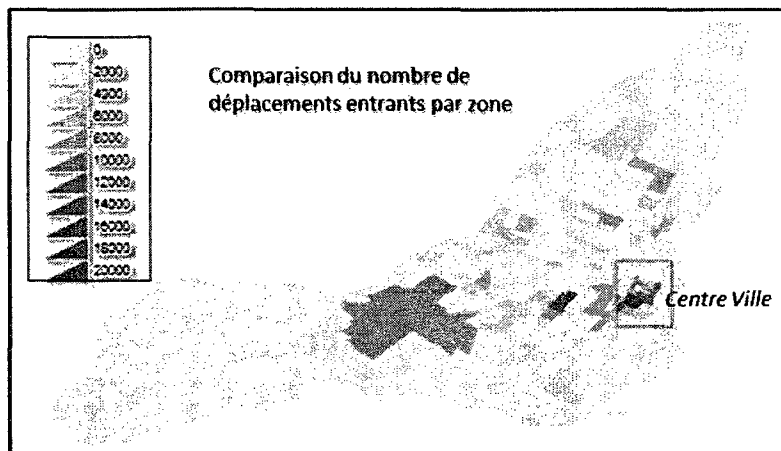
«  $\text{total\_hor}\sim 6$  » («  $\text{total\_hor}\sim 6$  » ) représente le nombre moyen de déplacements effectués par un habitant du secteur vers un secteur hors du centre ville (vers un secteur du centre ville).

Les données de recensement, de réseau TC et résultats de l'enquête OD ont ainsi pu être mises en commun pour la suite de l'étude.

#### **5.4.2 Étude des déplacements par rapport aux données de population**

Afin de se familiariser avec les données et les outils à disposition, nous avons analysé le comportement de l'utilisateur prenant le transport en commun à Montréal à partir de représentations graphiques des déplacements en fonction de leur profil (âge, sexe) et du motif du voyage.

*Problématique : Se familiariser avec les données et effectuer une interprétation par rapport aux nombres de déplacements, profil de l'utilisateur, motif.*

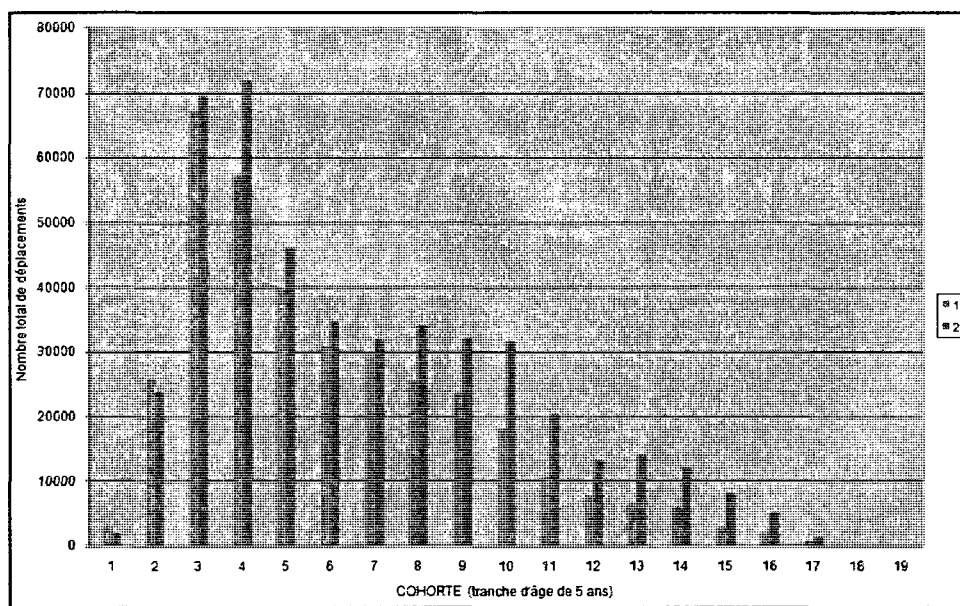


**Figure 5.3: Nombre de déplacements entrants par SR**

À titre d'exemple, la figure 5.3 montre les tendances au niveau de la destination des déplacements. On définit alors le centre ville (encadré bleu) comme principale zone d'attraction.

Notre analyse montre également que les femmes empruntent globalement plus le TC que les hommes à tout âge. La figure 5.4 met bien en évidence la différence en nombre de déplacements entre les deux sexes en fonction de leur classe d'âge. Même si la variation globale est similaire et que le maximum est atteint dans la même classe pour les deux, les femmes sont plus enclines à utiliser le transport en commun que les hommes.

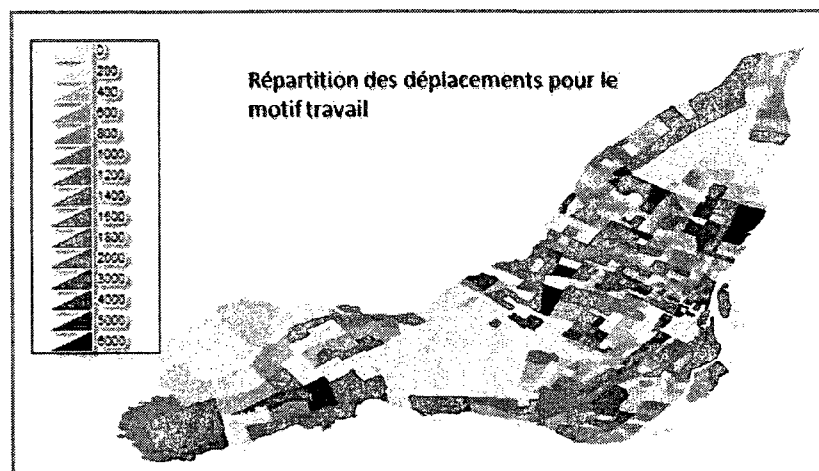




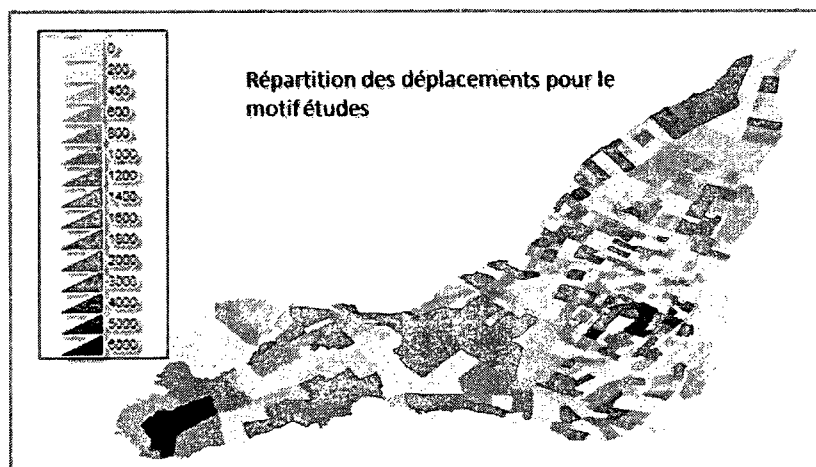
**Figure 5.4: Nombre de déplacements des usagers par rapport à leur sexe (homme:1, femme:2) et leur tranche d'âge**

En ce qui concerne les raisons des déplacements, les motifs étude et travail sont les plus caractéristiques de la pointe du matin. L'étude géographique à partir du logiciel FGIS nous donne une visualisation spatiale de la répartition des déplacements. Celui-ci nous a permis de localiser les valeurs des indicateurs. Ainsi, pour certaines variables connues pertinentes, il est intéressant de comparer leur répartition à différentes échelles de découpage par exemple.

Les figures 5.5 et 5.6 sont données à titre d'exemples. Les analyses ont pu être effectuées sur l'ensemble de la région de Montréal, néanmoins l'île de Montréal est notre centre d'intérêt.



**Figure 5.5: Répartition des déplacements entrants ayant pour motif le travail**



**Figure 5.6: Répartition des déplacements entrants ayant pour motif les études**

La Figure 5.5 décrit le nombre de déplacements attirés par les secteurs ayant pour motif le travail. On repère alors les zones industrielles de Montréal et le CBD (Center Business District) qui, évidemment, représentent les points de concentration des activités. La Figure 5.6 représente les déplacements pour motif études. Les zones d'études sont plus éparpillées

et les déplacements associés sont moindres par rapport au motif travail. On les retrouve proches du centre ville et dans les banlieues.

### **5.4.3 Étude des déplacements vers Centre Ville/Hors Centre Ville**

Comme préliminaire à la méthode hybride, il est intéressant d'étudier la différence entre le comportement des usagers allant au CV et celui de ceux n'y allant pas. A partir des données d'enquête OD, nous avons étudié les variations des déplacements par rapport aux caractéristiques de la population et au motif du déplacement (études-travail uniquement). Pour cela, nous avons séparé les données par rapport au centre ville.

*Problématique : Expliquer les différences entre les déplacements vers et hors CV en termes de nombre de déplacements, profil de l'utilisateur, motif.*

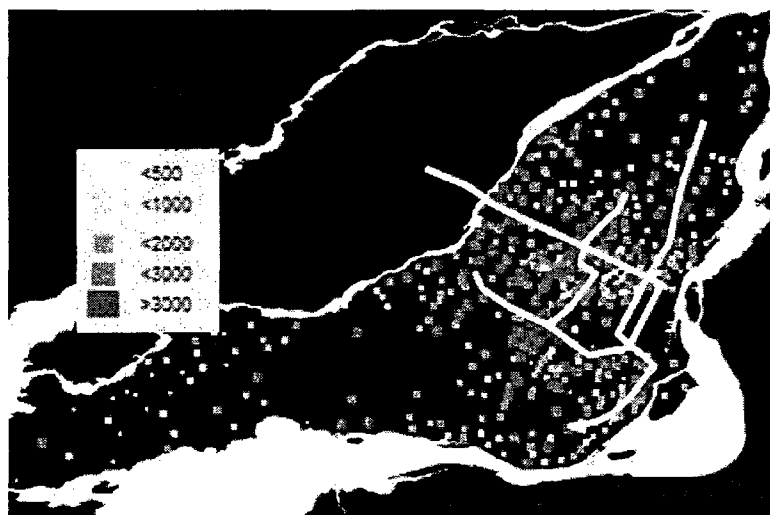
Pour ce faire, nous avons utilisé le logiciel FGIS pour représenter spatialement les données et Excel pour les graphiques pour chacun des cas suivants :

- totalité des déplacements (résultats ci-après);
- déplacements par sexe;
- déplacements par tranche d'âge (cohorte);
- déplacements par motif.

Pour chaque variable de déplacements, nous avons différencié les destinations CV et hors CV. La suite de cette sous-partie est consacrée aux résultats de l'expérimentation réalisée pour la totalité des déplacements.

Par ailleurs, la vision globale des indicateurs sous forme géographiée a notamment permis de mettre en évidence la trame des lignes de métro de Montréal. En effet, la majorité des

déplacements TC de la ville se font à proximité des lignes de métro et de train ou plus exactement dans les SR où se trouvent des stations.

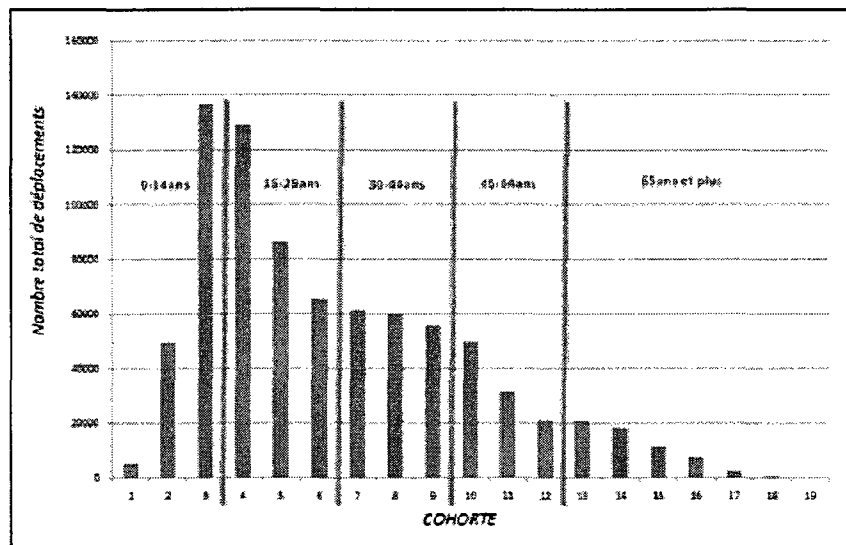


**Figure 5.7: Répartition des déplacements à destination Hors-CV de Montréal**

La figure 5.7 montre que les déplacements hors CV sont également concentrés proches des lignes de métro et les zones résidentielles. Le TC est majoritairement sollicité dans les quartiers Côtes des Neiges, Rosemont, Villeray et Plateau Mont-Royal.

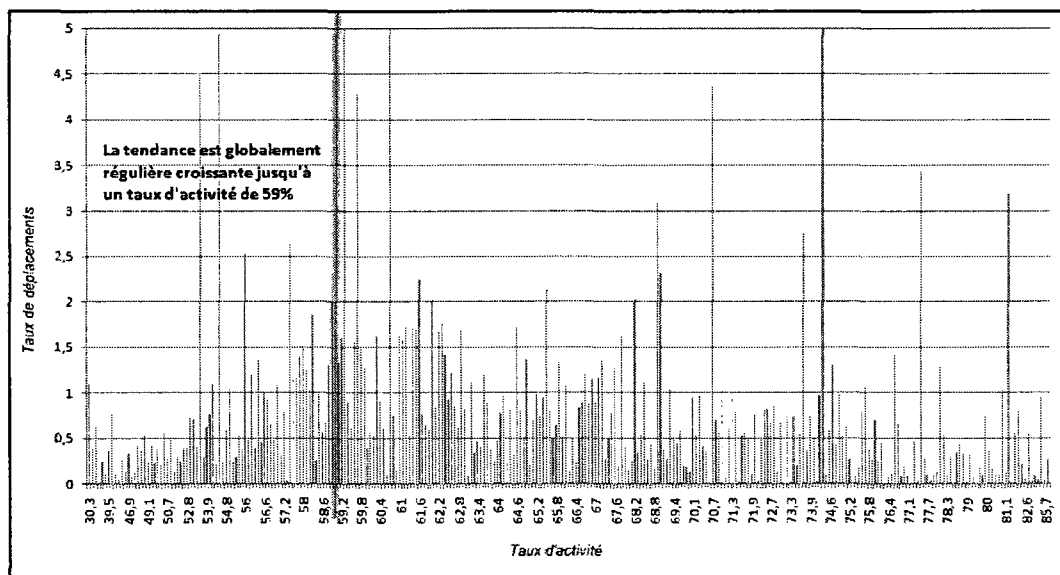
#### **5.4.4 Stratification des variables**

L'étude préliminaire consiste enfin à modifier les variables, c'est-à-dire identifier les « phases » de tendance qui constituent l'évolution de chaque variable en fonction des déplacements TC. L'étude graphique de l'évolution de chaque indicateur par rapport au nombre de déplacements nous a permis d'identifier des catégories de variables. Les figures 5.8 et 5.9 sont des exemples de stratification possible.



**Figure 5.8: Définition des groupes d'âge (selon les cohortes en abscisse) par rapport au nombre de déplacements (en ordonnée)**

Nous avons défini des groupes d'âge pour que l'étude des comportements soit moins complexe. La figure 5.8 fait clairement apparaître des tendances de déplacement par groupe cohorte, la définition de cinq groupes d'âge s'est alors faite intuitivement.



**Figure 5.9: Stratification de la variable taux d'activité par rapport au taux de déplacements**

La figure 5.9 présente le taux de déplacements en fonction du taux d'activité d'une zone. Jusqu'à environ 59% d'activité, la propension au déplacement grandit. Au-delà de ce seuil, il est difficile de distinguer les tendances. La variable du taux d'activité est transformée en deux variables catégoriques traduisant pour chaque zone si le taux d'activité est inférieur à 59% ou non.

Ainsi, pour chaque variable explicative, nous avons défini au plus 3 catégories (2 préférablement) pour lesquelles la variation du nombre de déplacements change significativement par rapport au reste. La majorité des variables ont été stratifiées mais il faut préciser que la stratification que nous présentons est pertinente pour ce cas particulier mais pourrait convenir à un autre territoire d'étude aux caractéristiques socio-économiques et démographiques similaires. La catégorisation traduit le comportement de l'utilisateur moyen

en fonction de sa situation socio-économique. Nous obtenons finalement de nouvelles variables que nous présentons dans le tableau 4 ci-après.

**Tableau 5.2: Variables explicatives de l'étude**

Variable	Description
txactiv0_59	taux d'activité du SR si <59% = 1 sinon 0
txactiv59	taux d'activité du SR si >=59% = 1 sinon 0
txchom0_8	taux de chômage du SR si <8% = 1 sinon 0
txchom8	taux de chômage du SR si >=8% = 1 sinon 0
revenu<30000	revenu moyen des ménages du SR si <30000 = 1 sinon 0
0_14	pourcentage population entre 0 et 14 ans
15_29	pourcentage population entre 15 et 29 ans
30_44	pourcentage population entre 30 et 44 ans
45_64	pourcentage population entre 45 et 64 ans
65	pourcentage population entre 65 ans et plus
fem_50	pourcentage de femmes si <50% = 1 sinon 0
fem_500	pourcentage de femmes si >=50% = 1 sinon 0
1p	pourcentage ménages à 1 personne
3p	pourcentage ménages à 3 personnes
4p	pourcentage ménages à 4 personnes et plus
dens_2500	densité de la population du SR si <2500 = 1 sinon 0
dens_25000	densité de la population du SR si >=2500 = 1 sinon 0
nbll_3	nombre de lignes de transport collectif différentes si <3 = 1 sinon 0 desservant le SR
nbll0_32	nombre de lignes si entre 10 et 32 = 1 sinon 0

**Tableau 5.2: Variables explicatives de l'étude (suite et fin)**

nb32	nombre de lignes si $\geq 32 = 1$ sinon 0
nba1_10	nombre d'arrêts du SR si $< 10 = 1$ sinon 0
nba10_32	nombre d'arrêts du SR si entre 10 et 32 = 1 sinon 0
nba32	nombre d'arrêts du SR si $\geq 32 = 1$ sinon 0
nbv0_46	nombre de voyages total (autobus et métro) qui traversent le SR si $< 46 = 1$ sinon 0
nbv46_132	nombre de voyages si entre 46 et 132 = 1 sinon 0
nbv132	nombre de voyages si $\geq 132 = 1$ sinon 0
nbmet_pie	nombre de voyages de métro qui traversent le secteur
dist_mét<350	distance à la station de métro ou de train la plus proche (si 0, il y a un métro ou un train dans le secteur) si $< 350m$
dist_mét>=350	distance à la station la plus proche si $\geq 350m$
superficie	superficie du SR
disten	distance en km

## 5.5 Synthèse

Finalement, l'étape de traitement de données a produit une série de variables catégorisées plus pertinentes par rapport au sujet de recherche. L'analyse préliminaire (graphique et géographique) permet de dresser un portrait du territoire par rapport à la population et aux tendances de déplacements (par âge, sexe, motif etc.). La stratification adapte les variables de façon à ce qu'elles soient propres à l'étude.

Il serait intéressant de réaliser cette étape pour d'autres territoires afin d'envisager une généralisation de la stratification pour l'estimation de la demande TC.



## CHAPITRE 6. GÉNÉRATION DES DÉPLACEMENTS

### 6.1 Étape 1 : Génération classique

À l'issue de la collecte des données et de l'analyse préliminaire des données, la procédure que nous suivons consiste à estimer les déplacements émis et attirés dans la région d'étude. Cette étape est exactement la même que pour la PSC.

Rappelons tout d'abord que la région a été spécifiée sous forme de zones pour identifier les déplacements de manière agrégée. Il a donc fallu choisir un système zonal à partir duquel tout le reste de l'étude se basera pour interpréter les résultats. Celui-ci est un des types de découpage administratif à disposition. Puis, le regroupement des variables explicatives associées à chaque zone a permis une analyse préliminaire en vue d'une stratification si des classes de variables ou de zones apparaissent ainsi que d'une étude des corrélations.

À partir des données adaptées à l'étude, nous estimons les coefficients du modèle de régression décrivant les variations des déplacements à partir des caractéristiques de chaque zone qui génère automatiquement la production et l'attraction de déplacements dans la région.

La forme générale du modèle est la suivante :

$$E_i = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n + e_i \quad (11)$$

$$A_i = \beta_1 X'_1 + \beta_2 X'_2 + \dots + \beta_m X'_m + e'_i \quad (12)$$

où :

$i$  est la zone

$n$  est le nombre de variables considérées pour modéliser les déplacements produits

$m$  est le nombre de variables considérées pour modéliser les déplacements attirés

$E_i$  est le nombre de déplacements par personne produits par la zone  $i$

$A_i$  est le nombre de déplacements par personne attirés par la zone  $i$

$X_1, \dots, X_n$  sont les variables indépendantes (impliquant la production de déplacements)

$X'_1 \dots X'_n$  sont les variables indépendantes (impliquant l'attraction de déplacements)  
 $\alpha_1 \dots \alpha_n$  et  $\beta_1 \dots \beta_n$  sont les coefficients de régression à estimer  
 $e_i$  et  $e'_i$  sont les résidus

### 6.1.1 Méthodes de sélection des variables

En statistique, il existe des algorithmes de sélection des variables explicatives les plus pertinentes afin d'optimiser le modèle de prédiction. Nous présentons ici trois de ces méthodes. Les résultats obtenus ne sont pas pris en compte tels quels pour le modèle final mais servent de guide pour le choix des meilleures variables. Les logiciels de statistiques sont dotés de ces algorithmes et peuvent les réaliser automatiquement.

#### Sélection avant (BACKWARD STEPWISE)

Cette méthode consiste à effectuer une régression sur l'ensemble du modèle (inclure l'ensemble des variables explicatives disponibles) et de sélectionner à chaque étape la variable avec la plus petite valeur de  $p$  (significativité la plus grande dans le modèle). Cette variable est intégrée au modèle optimal. La condition d'arrêt est la valeur maximale  $P$  que  $p$  ne doit pas dépasser. Lorsque toutes les variables restantes ont une valeur de  $p$  supérieur à  $P$ , le modèle optimal est obtenu. Ce modèle ne contient alors que des variables pertinentes. Il est possible de réaliser l'algorithme dans le sens inverse : on élimine une à une les variables qui ne sont pas suffisamment significatives.

#### Maximisation du $r^2$

Pour cette méthode, il faut estimer les modèles pour toutes les combinaisons de variables possibles. Pour  $p$  variables, il y a  $2^p$  modèles à réaliser. À l'issue des estimations, les modèles doivent être classés en fonction du coefficient de détermination  $R^2$  (de manière décroissante). À partir de la liste obtenue, le modèle « optimal » est celui avec une valeur de  $R^2$  convenable par rapport au maximum et dont le nombre de variables reste intéressant pour l'interprétation des résultats et la prévision.

### Coefficient de Mallows ( $c_q$ )

La méthode consiste à calculer les valeurs des coefficients de Mallows pour tous les modèles possibles comme précédemment et de ne garder que ceux pour lesquels les valeurs de  $C_q$  sont les plus petites. Les coefficients sont calculés à partir de l'erreur quadratique du modèle considéré. Il s'agit alors d'effectuer la régression sur les meilleurs modèles et de s'appuyer sur les valeurs du coefficient de détermination  $R^2$  et des tests de significativité pour sélectionner le meilleur modèle.

La sélection des variables explicatives est la partie critique de la génération des déplacements puisqu'elle définit le modèle. Bonnel (2001) a proposé un certain nombre de variables qui se retrouvent généralement dans la fonction de régression. Il est alors nécessaire de vérifier leur pertinence dans le contexte de l'étude et que les corrélations ne soient pas fortes entre elles. Le choix d'une variable doit être motivé par sa corrélation forte avec la variable à expliquer (nombre de déplacements) et par la faible corrélation avec les autres variables explicatives. Cette seconde condition n'est pas vérifiée dans les algorithmes de sélection, c'est pourquoi il est important d'effectuer des analyses complémentaires.

### **6.1.2 Génération des déplacements à partir d'un solveur**

L'étape de génération des déplacements permet d'estimer pour chaque zone  $i$  (groupement ou non) le nombre de déplacements entrants et sortants par personne (taux) en fonction des caractéristiques de celle-ci. Si celle-ci est effectuée à partir d'un solveur, les problèmes suivants doivent être formulés :

$$\begin{array}{l} \text{Min } \sum_{i=1}^n (E_i - P_i)^2 \\ E_i = \sum_{j=1}^m \alpha_j X_{ij} \\ E_i \geq 0 \end{array} \quad (13)$$

$$\begin{array}{l} \text{Min } \sum_{i=1}^n (A_i - O_i)^2 \\ A_i = \sum_{j=1}^l \alpha_j X_{ij} \\ A_i \geq 0 \end{array} \quad (14)$$

Où :

$P_i$  et  $O_i$  sont les quantités de déplacements de référence (que l'on veut approcher)  
 $n$  est le nombre de zones considérées (le plus agrégé, avec des groupements de zones)  
 $m$  est le nombre de variables explicatives intégrées pour prédire les déplacements émis  
 $l$  est le nombre de variables explicatives intégrées pour prédire les déplacements attirés  
 $X_{ij}$  sont les variables explicatives  
 $\alpha_j$  sont les coefficients à estimer

Le calcul des indicateurs de performance se fait alors manuellement. Par exemple, pour le coefficient de détermination on utilise la formule suivante :

$$R^2 = \frac{\sum (E_i - \bar{P})^2}{\sum (P_i - \bar{P})^2} \quad (15) \text{ où } \bar{P} \text{ est la moyenne des déplacements de référence.}$$

### 6.1.3 Quelques résultats d'expérimentation

Pour notre expérimentation, nous avons adapté et combiné les données de l'enquête OD 2003 et de recensement 2001 pour effectuer la modélisation des déplacements TC de l'île de Montréal. Nous avons alors obtenu un ensemble de données statistiques comprenant des variables continues mais aussi des variables catégoriques (1,0) qui rendent les données plus facilement modélisables. Nous donnons ici les résultats de la modélisation réalisée à partir de STATA et du solveur EXCEL.

Les variables classiques pour modéliser les déplacements que nous avons listées dans la section qui présente la PSC (cf. section 3.2.2) ont été utilisées.

#### **Modélisation des $E_i$ (déplacements sortants) pour la population totale tous âges et sexes confondus**

. regress txdepl txactiv0_59 txactiv59_ txchom8_ _4p_ _nbmet_pte nb110_32					
Source	SS	df	MS	Number of obs = 41	
Model	2.82308971	6	.470514952	F( 6, 34) =	74.83
Residual	.213773056	34	.006287443	Prob > F =	0.0000
Total	3.03686277	40	.075921569	R-squared =	0.9296
				Adj R-squared =	0.9172
				Root MSE =	.07929
txdepl	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
txactiv0_59	.3937641	.0913628	4.31	0.000	.2080926 .5794356
txactiv59_	.2906574	.0832195	3.49	0.001	.1215351 .4597798
txchom8_	-.1121264	.0373286	-3.00	0.005	-.1879872 -.0362656
_4p_	-.4581525	.1549686	-2.96	0.006	-.7730865 -.1432185
_nbmet_pte	.0236492	.0013371	17.69	0.000	.0209318 .0263666
nb110_32	.1998229	.0855561	2.34	0.026	.0259521 .3736937
_cons	-.0326208	.0942379	-0.35	0.731	-.2241352 .1588936

**Figure 6.1: Résultat de la génération des déplacements sortants pour le découpage principal de l'île de Montréal à partir de STATA avec application des méthodes de sélection de variables**

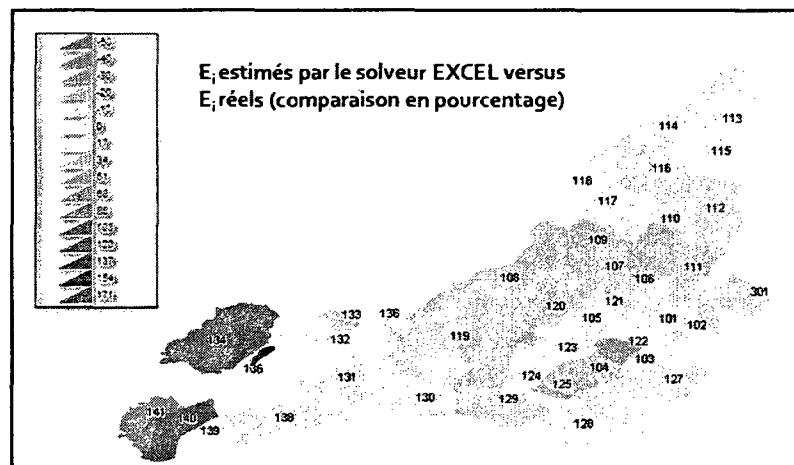
Les meilleurs résultats obtenus à partir de STATA après application des méthodes de sélection des variables pour l'ensemble des déplacements sortants (tout profil confondu) sont donnés dans la figure 6.1. Le nombre de passages du métro dans la zone est un moteur important de mobilité TC car il s'agit d'une des variables les plus significatives. Le taux d'activité est également important, les zones où il y a peu d'activités voient la propension des personnes à sortir de la zone plus importante. Enfin on remarque que le taux de chômage élevé de certaines zones diminue l'utilisation du TC : une zone où le taux est supérieur à 8% provoque une diminution de plus de 10% des déplacements TC.

$$\begin{aligned}
 E_i = & -0,01 * tx\_chom8 + + 0,04 * revenu0\_30000 + 0,38 * \%0\_14 + \\
 & 0,04 * \%15\_29 - 1,57 * \%30\_44 - 0,27 * \%45\_64 - 0,65 * \%65 + + \\
 & 0,68 * \%men1p + 1,32 * \%men3p - 0,33 * \%men4p + + 0,05 * nb110\_32 + \\
 & 0,002 * nbmet\_pte + 0,01 * distmetro350
 \end{aligned}
 \tag{16}$$

Les variables sélectionnées pour le modèle représenté par l'équation 16 sont issues de la liste des variables pertinentes pour la modélisation énoncées par Bonnel (cf. 3.2.2). Les

coefficients du modèle estimés par le solveur (qui est utilisé pour le modèle final) montrent que les proportions d'âge et de taille de ménage des zones sont des facteurs pertinents pour expliquer les variations de l'usage du TC. Les 15-29ans sont enclins à emprunter le TC comparés aux personnes plus âgées (sans doute dû à la possession d'un véhicule personnel) de même que les ménages de petite taille.

Finalement, en comparaison avec les valeurs de référence, il est possible de modéliser géographiquement la performance des modèles obtenus. Nous présentons ci-dessous la carte de comparaison du pourcentage de différence entre les déplacements réels et ceux estimés par le solveur. Les estimations sont généralement à moins de 15% autour de la réalité avec une moyenne générale de 10% au dessus de la réalité. Les différences les plus importantes concernent les zones extrêmes de l'île.



**Figure 6.2: Comparaison des déplacements sortants réels et ceux estimés par le solveur EXCEL**

**Modélisation des  $A_i$  (déplacements attirés) pour la population totale tous âges et sexes confondus**

regress ai_réel _txchom8_ _revenu0_30000 _65_ _3p _4p_ _dens_2500 _nb11					
Source	SS	df	MS	Number of obs = 42	
Model	.000168075	10	.000016807	F( 10, 31) =	15.21
Residual	.000034256	31	1.1050e-06	Prob > F =	0.0000
Total	.000202331	41	4.9349e-06	R-squared =	0.8307
				Adj R-squared =	0.7761
				Root MSE =	.00105
ai_réel	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
_txchom8_	-.0019392	.0004849	-4.00	0.000	-.0029282 -.0009501
_reven~30000	-.0009494	.0005037	-1.88	0.069	-.0019767 .0000778
_65_	.0112868	.0045478	2.48	0.019	.0020115 .020562
_3p	.0428447	.0116464	3.68	0.001	.0190916 .0665977
_4p_	-.0146956	.0041026	-3.58	0.001	-.0230628 -.0063284
_dens_2500	.00199	.0005023	3.96	0.000	.0009656 .0030143
_nb11_3	-.0009291	.0004222	-2.20	0.035	-.0017902 -.0000679
_nb110_3_2	.0059138	.0012158	4.86	0.000	.0034342 .0083935
_nbmet_pte	.0001471	.0000197	7.45	0.000	.0001068 .0001873
_superficie	-.2592941	.0617809	-4.20	0.000	-.3852971 -.1332912
_cons	.0005996	.0016637	0.36	0.721	-.0027935 .0039927

**Figure 6.3: Résultats de la génération des déplacements attirés pour le découpage principal de l'île de Montréal à partir de STATA avec application des méthodes de sélection de variables**

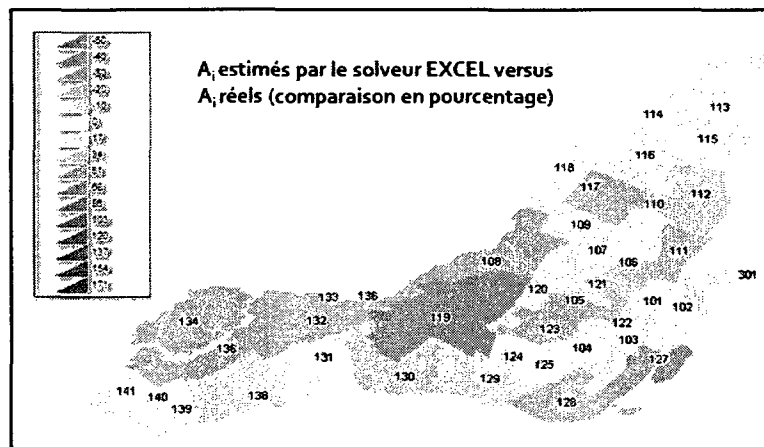
Selon le modèle proposé de la figure 6.3, pour tous âges et sexes confondus, le nombre de passages du métro en période de pointe, le nombre de lignes de bus passant par la zone et le taux élevé de chômage dans la zone de destination sont déterminants dans le choix d'emprunter le TC pour effectuer un déplacement. On en déduit qu'il s'agit du dynamisme (pourcentage de personnes qui possèdent un emploi) d'une zone qui attire plus ou moins les usagers. Le profil de l'utilisateur qui emprunte le TC pour rentrer chez lui serait une personne active et dont la taille du ménage est inférieure à 4 personnes. Sa zone de résidence devrait être bien desservie en TC (ligne de bus) et de préférence posséder au moins une station de métro.

$$\begin{aligned}
A_i = & 2,5E-4 * tx\_activ59 + - 0,002 * tx\_chom8 + - 0,001 * revenu0\_30000 - \\
& 0,002 * \%0\_14 + 0,02 * \%15\_29 - 0,02 * \%30\_44 + 0,004 * \%45\_64 + \\
& 0,007 * \%65 + - 8E-5 * \%men1p + 0,06 * \%men3p - 0,02 * \%men4p + + \\
& 0,003 * dens0\_2500 - 0,001 * nbl1\_3 + 0,005 * nbl10\_32 + 0,001 * nbla0\_32 + \\
& 6E-4 * nba35 + - 5E-4 * nbv46\_132 + 4,5E-4 * nbv132 + - \\
& 1E-4 * nbmet\_pte - 6E-4 * dist\_metro350 + - 0,23 * superficie
\end{aligned} \tag{17}$$

L'équation 17 montre que la modélisation du solveur met en évidence l'importance du profil de la population (âges et ménages) avec des coefficients de régression plus élevés. Les variables les plus influentes ici sont : le taux de chômage élevé, la proportion des 15-29ans (jeunes) et des 30-44ans (jeunes salariés), la taille du ménage, la densité de population, le nombre de lignes de bus traversant la zone et la superficie. On suppose ici que la superficie est pertinente par rapport à la définition du découpage : la superficie de la zone n'est pas corrélée à la population (qui est une variable constante) donc une zone plus grande n'engendrera pas forcément plus de déplacements TC. De plus, les superficies élevées concernent les banlieues éloignées où les personnes se déplacent moins en TC.

La carte ci-après compare les résultats obtenus de la génération des déplacements attirés par le solveur EXCEL (qui ont été présentés plus haut) par rapport aux  $A_i$  réels. On constate que les résultats du solveur sont très bons avec une moyenne de 3% au dessus de la réalité.





**Figure 6.4: Comparaison des déplacements attirés réels et estimés par le solveur EXCEL**

## 6.2 Étape 2 : Génération/Distribution

Cette étape se concentre sur les zones regroupées où les  $E_i$  et  $A_i$  doivent être distribués. L'étude des corrélations entre les données caractérisant les zones et les déplacements a été la base de ce modèle de distribution. Nous avons simplement recensé les variables (ou plutôt la variable) dont les variations sont les plus similaires à celle des  $E_i$  et  $A_i$ .

Il s'agit de réaliser à nouveau la génération des déplacements sous la forme d'un problème à contraintes linéaires c'est-à-dire d'effectuer à nouveau une régression linéaire multiple uniquement sur les zones du regroupement tout en imposant le nombre total de déplacements sortants  $E_i$  ou déplacements entrants  $A_i$  qui ont été calculés à l'étape précédente. Cette méthode implique à nouveau préalablement la sélection des variables indépendantes pertinentes pour la régression. Les variables pertinentes sont alors utilisées pour la méthode hybride où de nouveaux coefficients sont estimés lors de la résolution du problème de minimisation de la somme des différences au carré sous la contrainte du  $E_i$  ou

$A_i$  total. Ce modèle dépend complètement de la qualité des données et de « l'interprétabilité » du comportement des usagers avec celles-ci.

On obtient alors les problèmes suivants pour une zone  $i$  découpée en  $m$  (sous) zones:

$$\begin{aligned} & \text{Min } \sum_{j=1}^m (E_{ij} - P_{ij})^2 \\ & E_{ij} = \sum_{k=1}^k \alpha_k X_{ijk} \\ & \sum_{j=1}^m E_{ij} = E_i \\ & E_{ij} \geq 0 \end{aligned} \quad (18)$$

$$\begin{aligned} & \text{Min } \sum_{j=1}^m (A_{ij} - O_{ij})^2 \\ & A_{ij} = \sum_{k=1}^l \alpha_k X_{ijk} \\ & \sum_{j=1}^m A_{ij} = A_i \\ & A_{ij} \geq 0 \end{aligned} \quad (19)$$

Où :

$P_{ij}$  et  $O_{ij}$  sont les quantités de déplacements de référence (que l'on veut approcher)

$m$  est le nombre de zones du groupement  $i$

$k$  est le nombre de variables explicatives intégrées pour prédire les déplacements produits

$l$  est le nombre de variables explicatives intégrées pour prédire les déplacements attirés

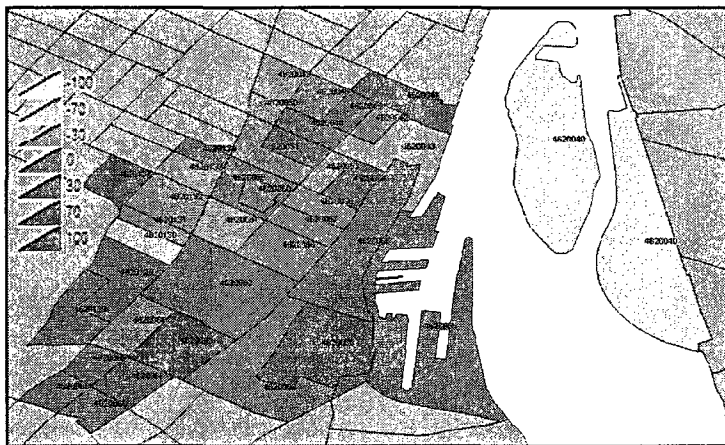
$X_{ijk}$  sont les variables explicatives

$\alpha_j$  sont les coefficients à estimer

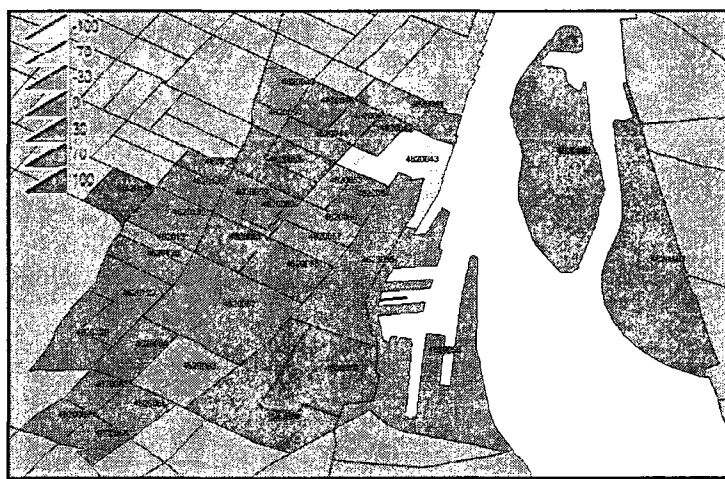
### 6.2.1 Quelques résultats d'expérimentation

Les résultats du solveur EXCEL de l'étape précédente ont été directement compilés pour réaliser la redistribution des déplacements. Afin d'optimiser l'estimation des coefficients, nous avons intégré l'ensemble des variables explicatives disponibles au modèle. Nous avons également effectué une interprétation au niveau géographique de la performance du modèle en comparaison avec les résultats obtenus de la PSC appliquée aux SR pour l'étape de génération.

Les cartes suivantes concernent uniquement les déplacements produits et utilisent la même échelle afin d'assurer la bonne comparaison des résultats. Les résultats sont équivalents pour les déplacements entrants.



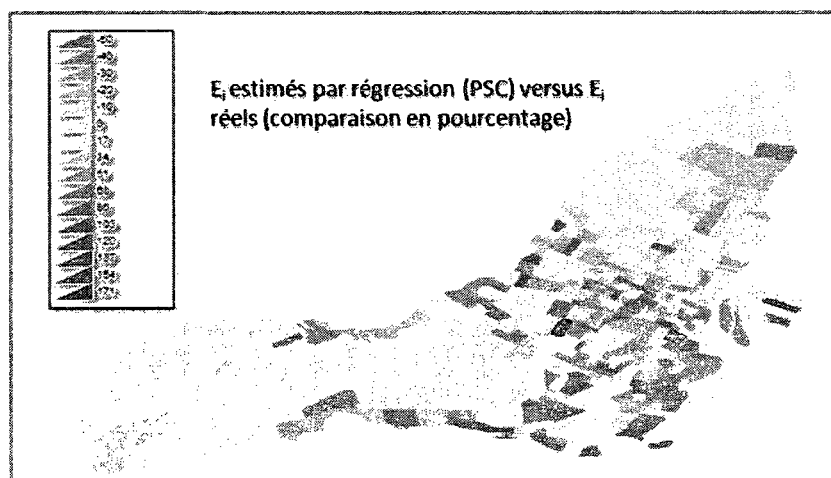
**Figure 6.5: Étape de génération-distribution, comparaison en pourcentage des déplacements produits ( $E_i$ ) réels par rapport à ceux estimés par la méthode hybride**



**Figure 6.6: Étape de génération-distribution, comparaison en pourcentage des déplacements émis ( $E_i$ ) réels par rapport à ceux estimés par la PSC**

Il résulte de l'expérimentation que le processus de génération des déplacements du modèle hybride donne de meilleurs résultats que le modèle classique. Les estimations présentées pour la PSC sont issues de la régression des variables explicatives pour le découpage en SR pour toute l'île de Montréal. Les variations du nombre de déplacements entre les différents secteurs sont donc plus difficilement explicables et les particularités du CV n'auront pas pu être mises en évidence.

La figure suivante traduit la performance de la génération classique sur les SR.



**Figure 6.7: Comparaison des déplacements produits réels et estimés par génération classique**

En comparaison avec la figure 6.4, les déplacements des secteurs éloignés sont mieux estimés contrairement aux secteurs proches du CV où l'on remarque des différences plus importantes pour la méthode classique par rapport à la méthode hybride. Les résultats de notre méthode ont un caractère plus homogène puisque le découpage considéré est moins fin (secteurs municipaux et non secteurs de recensement).

Le comportement de l'utilisateur sera plus difficilement interprétable par la génération classique puisque les décalages par rapport à la réalité ne sont pas « homogènes » c'est-à-

dire que les coefficients obtenus ne traduisent pas complètement la réalité pour l'ensemble des secteurs mais seulement pour une partie de ceux-ci. D'après la figure 6.7, le modèle obtenu est intéressant pour les banlieues éloignées du CV.

### 6.3 Synthèse

La génération des déplacements par la méthode hybride présente de meilleurs résultats que la méthode classique. Les coefficients sont estimés pour des systèmes de découpage adaptés : soit tout le territoire est considéré avec des zones larges, soit un petit ensemble de sous-zones sont fixées et correspondent à un profil particulier d'usagers. L'interprétation adéquate des résultats devient plus facile et la démarche associée à l'étape permet d'obtenir des données cohérentes à différentes échelles.

L'utilisation du solveur Excel est préconisée dans le sens où il s'agit de l'outil qui supporte les autres étapes du processus. L'application de cette étape sur le logiciel STATA a essentiellement permis d'évaluer la qualité du modèle au niveau statistiques plus rapidement.

## CHAPITRE 7. DISTRIBUTION DES DÉPLACEMENTS

La troisième étape de la procédure proposée consiste à remplir les matrices Origine-Destination contenant le nombre de déplacements effectués entre chaque zone à partir des résultats de la génération. Après avoir généré la demande, il est nécessaire de répartir les déplacements en fonction des secteurs accessibles par rapport à leur attractivité et au profil de la population. Nous avons choisi le modèle gravitaire utilisant la distance comme fonction d'impédance pour effectuer cette étape.

### 7.1 Première distribution : le modèle le plus agrégé

La méthode utilisée pour cette première étape est la même que celle présentée dans le chapitre précédent. Il s'agit de remplir la matrice à partir d'une formule de base que l'on modifie à chaque itération pour améliorer l'approximation par rapport à la matrice de référence.

#### 7.1.1 Détermination de la fonction d'impédance

La seule donnée de coût calculable à partir des données et des moyens techniques disponibles est la distance de déplacement. La distance entre deux zones données se calcule à partir de leurs centroïdes. Il est possible de calculer différents types de distance :

- la distance « à vol d'oiseau » qui se calcul directement à partir des coordonnées des centroïdes;
- la distance routière optimale, calculée à partir du logiciel SIG Microsoft Mappoint (route optimale entre deux points donnés);
- la distance moyenne en TC, la plus pertinente dans notre cas.

#### 7.1.2 Modèle gravitaire

Deux formulations du modèle gravitaire seront testées dans notre étude :

Le modèle avec la fonction puissance avec  $k=2$

$$T_{ij} = \frac{\alpha_i E_i \beta_j A_j}{d_{ij}^2} \quad (20)$$

où :

$T_{ij}$  est le nombre de déplacements entre la zone  $i$  et la zone  $j$

$\alpha_i$  et  $\beta_j$  sont des coefficients interdépendants mis à jour à chaque itération qui traduisent le respect des contraintes aux marges tels que  $\alpha_i = \frac{1}{\sum_{j=1}^n \frac{\beta_j A_j}{d_{ij}^2}}$  (21) et que

$$\beta_j = \frac{1}{\sum_{i=1}^n \frac{\alpha_i E_i}{d_{ij}^2}} \quad (22)$$

$E_i$  le nombre total de déplacements générés par la zone  $i$

$A_j$  le nombre total de déplacements attirés par la zone  $j$

$d_{ij}$  la distance entre les deux zones ( $f_{ij} = \frac{1}{d_{ij}^2}$ ) (23))

Le modèle avec la fonction exponentielle où  $k=2$

$$T_{ij} = \frac{\alpha_i E_i \beta_j A_j}{e^{2d_{ij}}} \quad (24)$$

où :

$T_{ij}$  est le nombre de déplacements entre la zone  $i$  et la zone  $j$

$\alpha_i$  et  $\beta_j$  sont des coefficients interdépendants mis à jour à chaque itération qui traduisent le respect des contraintes aux marges tels que  $\alpha_i = \frac{1}{\sum_{j=1}^n \frac{\beta_j A_j}{e^{2d_{ij}}}}$  (25) et que

$$\beta_j = \frac{1}{\sum_{i=1}^n \frac{\alpha_i E_i}{e^{2d_{ij}}}} \quad (26);$$

$E_i$  le nombre total de déplacements générés par la zone  $i$

$A_j$  le nombre total de déplacements attirés par la zone  $j$

$d_{ij}$  la distance entre les deux zones ( $f_{ij} = \frac{1}{e^{2d_{ij}}}$  (27))

### 7.1.3 Calage du modèle

Le calage (ou calibrage) du modèle se fait à partir de l'ajustement des coefficients  $\alpha_i$  et  $\beta_j$ . Après la génération des matrices OD avec  $\alpha_i = 1$  et  $\beta_j = 1$ , pour toute paire  $(i,j)$ , nous avons optimisé les valeurs des déplacements estimés par rapport aux déplacements réels (matrice de référence) en ajustant les  $\alpha_i$  et  $\beta_j$ . Nous imposons que les valeurs des coefficients varient de 0 à 5 (échelle de calibrage). Le calcul de ces valeurs permet alors d'améliorer significativement le modèle puisqu'il dépendra complètement des caractéristiques de la région d'étude.

La condition d'arrêt des itérations sera la minimisation de la somme des différences des écarts au carré entre les résultats et les valeurs de la matrice de référence.

### 7.1.4 Modèle gravitaire en utilisant le solveur

La réalisation de la matrice OD a impliqué l'utilisation du solveur EXCEL. Cela nous permet de manipuler directement les résultats de la génération sans avoir à changer d'outil.

En utilisant le solveur, le problème à codifier est le suivant :

$$\begin{aligned}
 & \text{Min } \sum_{i=1}^n \sum_{j=1}^n (T_{ij} - D_{ij})^2 \\
 & T_{ij} = \frac{\alpha_i E_i \beta_j A_j}{d_{ij}^2} \text{ ou } T_{ij} = \frac{\alpha_i E_i \beta_j A_j}{e^{2d_{ij}}} \\
 & T_{ij} \geq 0 \\
 & 10 \geq \alpha_i \geq 0 \\
 & 10 \geq \beta_j \geq 0
 \end{aligned} \tag{28}$$

Où :



$n$  est le nombre de zones du territoire d'étude ( $n \leq 200$ )

$T_{ij}$  est le nombre de déplacements estimés de la zone  $i$  à la zone  $j$

$\alpha_i$  et  $\beta_j$  sont les coefficients à estimer

$E_i$  et  $A_j$  sont les déplacements émis par la zone  $i$  et attirés par la zone  $j$  qui ont été estimés dans l'étape précédente

$d_{ij}$  est la distance entre la zone  $i$  et la zone  $j$  (calculée par Mappoint)

$D_{ij}$  est le nombre de déplacements réels donnés par la matrice de référence.

Lors du lancement du solveur, on pose les valeurs initiales des coefficients : pour tout  $i$  et  $j$ ,  $\alpha_i = 1$  et  $\beta_j = 1$ .

### 7.1.5 Indicateurs de reproductibilité

Afin de démontrer la fiabilité du modèle, nous comparons les variations des coefficients avec les attributs des zones (population, réseau de transport). L'interprétation des indicateurs statistiques classiques (cf. 4.5) nous permet d'évaluer la significativité des coefficients estimés par rapport aux caractéristiques effectives de la région d'étude. Ainsi, plus les valeurs des  $\alpha_i$  et  $\beta_j$  sont pertinentes, plus le modèle est fiable et reproductible puisqu'il ne détermine pas les déplacements au hasard.

### 7.1.6 Quelques résultats d'expérimentation

À partir des nombres de déplacements obtenus à l'étape de génération, nous avons pu remplir les marges de la matrice OD. Après calcul des distances routières optimales (non TC) entre chaque paire de zones avec MapPoint et montage de la matrice de référence avec les données issues de l'enquête OD, nous avons pu réaliser la distribution des déplacements (affectation du nombre de déplacements effectués entre toutes les paires de zones).

L'optimisation des résultats peut être plus ou moins longue en fonction du solveur. La minimisation de la formule des moindres carrés consiste à trouver la meilleure combinaison des  $\alpha_i$  et  $\beta_j$ . Par ailleurs, plus la fonction d'impédance est complexe, plus le temps de

résolution sera long. Nous avons donc recensé deux types de résultats : les résultats après un seul lancement du solveur (optimisation à partir de  $\alpha_i=1$  et  $\beta_j=1$ ) et, ceux après une série de lancements (où l'on garde les valeurs optimisées des  $\alpha_i$  et  $\beta_j$  à chaque itération).

Au final, nous obtenons pour la population totale, tous âges et sexes confondus, le nombre de déplacements réalisés par les résidents pour chaque couple de zones. La première validation des résultats se fait par une comparaison statistique simple entre la matrice de référence et la matrice OD. Dans notre cas, la matrice de corrélation (coefficient supérieur à 0,7) traduit des corrélations fortes entre les résultats des trois distributions (celle de référence, celle construite avec la fonction puissance et celle construite avec la fonction exponentielle), ce qui montre que le processus a été réalisé correctement.

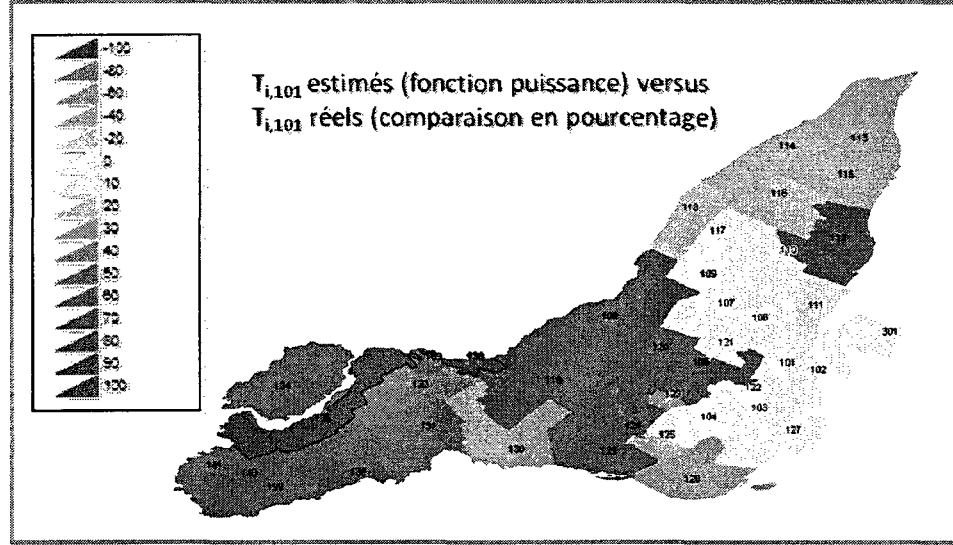
L'analyse statistique des données de la matrice de référence a mis en évidence quelques zones où la distribution est statistiquement explicable à partir des données socio-économiques et TC dont nous disposons. Le tableau ci-après présente une comparaison avec les variables explicatives les plus pertinentes pour les matrices OD de l'expérimentation :

**Tableau 7.1: Comparaison des variables pertinentes expliquant la distribution des déplacements attirés pour les secteurs de destination 101, 102 et 127**

	Matrice de référence		Matrice OD formule puissance		Matrice OD formule exponentielle	
	Variable la plus significative de la régression*	Variable avec la corrélation la plus forte**	Variable la plus significative de la régression*	Variable avec la corrélation la plus forte**	Variable la plus significative de la régression*	Variable avec la corrélation la plus forte**
101	nbmet_pte	%menlp	nbmet_pte	nbmet_pte	nbmet_pte	nbmet_pte
102	nbv0_46	%15_29ans	nbv0_46	%15_29ans	nbv0_46	%15_29ans
127	nbl10_32	nbl10_32	nbl10_32	nbl1_3	nbl_10_32	nbl10_32
<p>* Variable la plus influente dans le modèle de régression qui décrit la distribution des déplacements entrants dans la zone (101, 102 ou 127) en fonction des données socio-économiques des zones de départ.</p> <p>** Variable socio-économique la plus fortement corrélée à la distribution des déplacements entrants dans la zone (101, 102 ou 127).</p>						

Les résultats statistiques montrent que les distributions des déplacements sont pertinentes par rapport à la matrice de référence. Les variations s'appuient sur les mêmes variables socio-économiques et les distributions sont fortement corrélées.

Néanmoins, la comparaison des nombres de déplacements a montré les lacunes du solveur qui n'arrive pas à approximer correctement les données. Comme le présente la représentation géographique des différences entre les résultats et la réalité ci-après, les estimations ne sont pas très proches de la matrice de référence (plus de 50% d'écart). Ceci est dû au fait que le solveur n'est pas suffisamment performant pour résoudre un tel problème à la première itération. Les  $\alpha_i$  et  $\beta_j$  de départ sont déterminants. Ainsi, afin d'optimiser les résultats, il est nécessaire de relancer le solveur.



**Figure 7.1: Comparaison des déplacements produits se destinant à la zone 101 estimés avec ceux de la matrice de référence**

## 7.2 Distribution par désagrégation : affinage

Pour chaque paire de zones (i,j) pour laquelle il existe un groupement de zones, on crée une sous-matrice que l'on remplit comme la matrice primaire et pour laquelle on ajoute la contrainte du nombre total de déplacements qui vient d'être estimé.

$$\begin{aligned}
 & \text{Min } \sum_{a=1}^l \sum_{b=1}^m (T_{abij} - D_{abij})^2 \\
 & T_{abij} = \frac{\alpha_{ai} E_{ai} \beta_{bj} A_{bj}}{d_{abij}^2} \text{ ou } T_{abij} = \frac{\alpha_{ai} E_{ai} \beta_{bj} A_{bj}}{e^{2d_{abij}}} \\
 & \sum_{a=1}^l \sum_{b=1}^m T_{abij} = T_{ij} \\
 & T_{abij} \geq 0 \\
 & 10 \geq \alpha_{ai} \geq 0 \\
 & 10 \geq \beta_{bj} \geq 0
 \end{aligned} \tag{29}$$

Où :

$T_{ij}$  est le nombre total de déplacements pour la paire OD (i,j) où la zone i ou la zone j ou les deux sont des groupements de zones

$T_{abij}$  est le nombre total de déplacements pour la paire OD (a,b) issus de la paire de groupements (i,j)

$l$  est le nombre de (sous) zones pour la zone i

$m$  est le nombre de (sous) zones pour la zone j

$\alpha_{ai}$  et  $\beta_{bj}$  sont les coefficients à estimer

$E_{ai}$  et  $A_{bj}$  sont les déplacements émis par la zone a et attirés par la zone b de la paire de groupements (i,j) qui ont été estimés dans l'étape précédente

$d_{abij}$  est la distance entre la zone a et la zone b (calculée par Mappoint) de la paire de groupements (i,j)

$D_{abij}$  est le nombre de déplacements réels donnés par la matrice de référence

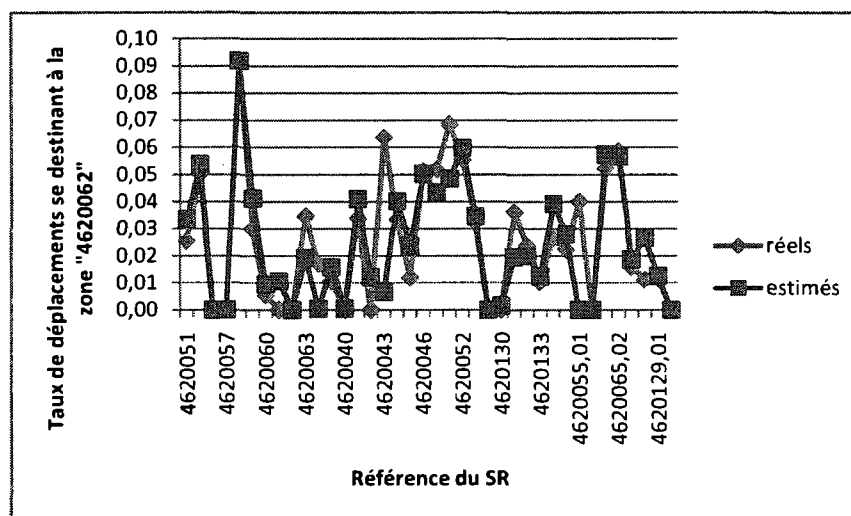
Cette étape suppose donc qu'un groupement de zones possède un nombre maximal possible de zones qui correspond à la limite de l'outil. Il est donc nécessaire que le nombre total de coefficients à estimer soit inférieur à 200 soit :  $l+m \leq 200$ .

### 7.2.1 Quelques résultats d'expérimentation

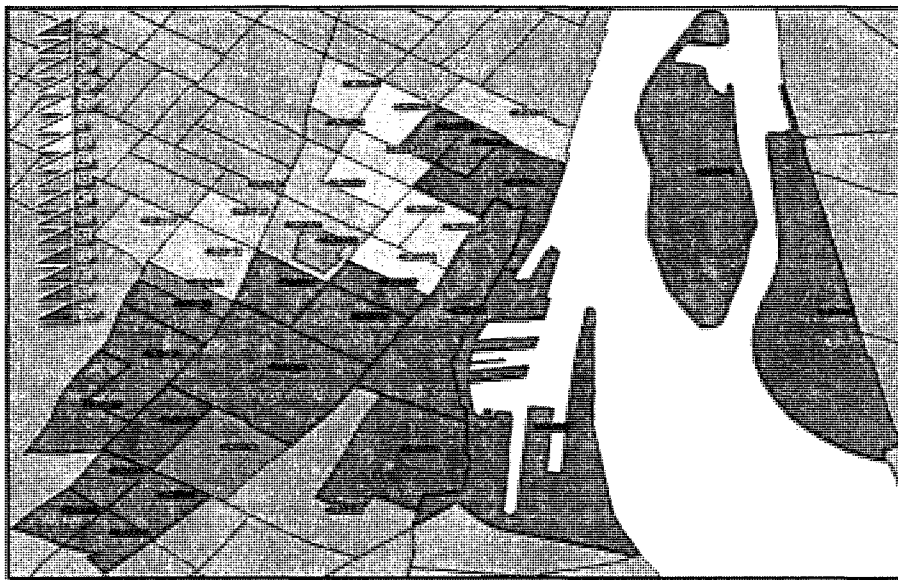
De même, nous avons réalisé le remplissage de la sous-matrice OD à partir de la matrice « primaire » et de la matrice de référence. Le principe de validation reste identique à celui de la distribution précédente. Cependant, la comparaison des résultats de la méthode hybride avec la procédure classique avec pour support le solveur Excel est impossible à cause de la limite en nombre de variables. Comme expliqué précédemment, l'île de Montréal est composée de plus de 500 SR, ce qui impose l'estimation de 1000 coefficients, ce qui est au-delà des capacités du solveur.

Par ailleurs, la résolution par le solveur devient plus complexe car celui-ci doit estimer les coefficients en fonction de la matrice « primaire ». Le temps de résolution et l'optimisation des résultats sont donc plus longs.

Nous présentons ici les résultats associés au secteur « 4620062 » (contient la Place Ville Marie et les principaux commerces du CV) pour la distribution avec la fonction puissance. La comparaison directe des quantités de déplacements estimés par rapport à la matrice de référence donne un coefficient de corrélation supérieur à 0.8 ; les variations calculées sont fortement similaires à la réalité. Le graphique suivant traduit effectivement la performance de la distribution.



**Figure 7.2: Comparaison en pourcentage des déplacements entrants dans la zone "4620062" estimés par rapport à la matrice de référence**



**Figure 7.3: Comparaison en pourcentage des déplacements entrants dans la zone "4620062" estimés par rapport à la matrice de référence**

La dernière figure indique une sous-estimation des déplacements par le solveur pour la plupart des zones du CV. Les zones en vert représentent des parties singulières du CV : centre d'affaires, parc d'attraction, ce qui explique la difficulté du solveur à mettre en place un coefficient de déplacement. Ici, les résultats auraient été mieux estimés en utilisant des coefficients de type  $k_{ij}$  (pour chaque paire de zone  $(i,j)$ ) puisqu'ils permettraient de distinguer les comportements des usagers par rapport à chaque zone et non pour tout le territoire.

### 7.3 Validation de la matrice complète

Lorsque chaque sous-matrice a été estimée, la matrice complète peut être mise en place. Dans notre expérimentation, une seule sous-matrice a été estimée (pour le CV). L'intégration de la sous-matrice se fait directement puisque les valeurs des déplacements sont en cohésion. Cependant, les parties non estimées de la matrice (les sous-zones par

rapport aux zones « primaires ») seront calculées directement à partir de la formule de distribution.

La validation s'effectue de nouveau par rapport à la matrice de référence et à partir de l'analyse comparative avec les variables de l'étude (données de population, socio-économiques, de réseau). L'étude des distributions pour les SM puis pour les SR du CV ayant été réalisées précédemment, cette partie consiste à analyser les déplacements vers le CV et à partir du CV vers les SM (hors CV).

Nous avons tout d'abord réalisé une analyse statistique pour mettre en évidence les relations existantes ou non entre les déplacements estimés et réels ainsi que les variables du recensement et du réseau : corrélation, définition des variables significatives. Il en résulte globalement que les déplacements à destination du CV ont été beaucoup mieux estimés que ceux sortant du CV et à destination hors CV.

Les résidents du CV ont un profil particulier et leurs déplacements vers des zones hors du CV ont des motifs particuliers : on suppose que les travailleurs du CV ont leur emploi au CV ainsi que les étudiants y ont leur lieu d'études. Les déplacements vers l'extérieur concerneront les motifs loisirs, magasinage et autres. Ces motifs ne sont pas significatifs en période de pointe, ce qui explique la difficulté du solveur à estimer ces déplacements particuliers. Les résultats obtenus ne présentent aucune corrélation significative avec la réalité ou les données socio-économiques.

Les résidents hors CV se déplaçant vers le CV ont un profil plus facile à définir dans le sens où le CV de Montréal est un noyau pour le travail, les études mais aussi le magasinage soit trois motifs importants considérés dans le modèle. Ainsi, des corrélations fortes c'est-à-dire des tendances similaires ont été détectées tant avec les déplacements de la matrice de référence que les variables explicatives de l'étude.



**Tableau 7.2: Matrice de corrélations entre les déplacements à destination de certaines zones du CV et les données de population et du réseau pour l'île de Montréal**

	4620040	4620041	4620049	4620051	4620057	4620059
txchom8	0.3365	0.3066	0.2363	0.2230	0.3627	0.2803
reven~30000	0.5196	0.4591	0.4316	0.4468	0.4875	0.4341
0_14	-0.5036	-0.5214	-0.5354	-0.5731	-0.5296	-0.5777
15_29	0.5764	0.6443	0.6857	0.7236	0.6140	0.7145
30_44	0.3464	0.3427	0.3837	0.4615	0.3981	0.4409
45_64	-0.3753	-0.3978	-0.4038	-0.4421	-0.4391	-0.4532
65	-0.1235	-0.1597	-0.1837	-0.2095	-0.1704	-0.1993
fem_50	0.5065	0.6311	0.6972	0.6446	0.4112	0.5769
1P	0.4963	0.5149	0.5155	0.5490	0.5460	0.5658
3P	-0.3766	-0.4650	-0.4923	-0.5298	-0.4733	-0.5362
4P	-0.5249	-0.5182	-0.5050	-0.5364	-0.5640	-0.5586
nbl1_3	0.3300	0.3743	0.3204	0.2423	0.3314	0.2614
nbal_10	0.4471	0.5637	0.5985	0.6576	0.5931	0.6418
nbmet_pte	0.1761	0.2487	0.2156	0.2531	0.3773	0.3014
distcv	-0.3297	-0.3401	-0.3243	-0.3812	-0.4454	-0.4243

Le tableau 6 compare les coefficients de corrélation entre les déplacements entrant dans certaines zones du CV et les variables socio-économiques les plus pertinentes de l'étude. Les valeurs sont assez similaires pour toutes les zones et traduisent bien le profil de l'utilisateur qui se destine au CV. Il s'agit généralement des 15-29 ans, vivant seuls et préférentiellement des femmes. La zone de résidence doit être bien desservie en TC (nombre d'arrêts, de métro). Par contre, la distance au CV ne varie pas de la même manière que l'utilisation du TC, ce qui signifie qu'il ne suffit pas à expliquer son usage.

## 7.4 Synthèse

La résolution du problème de distribution a pu être réalisée aisément par le solveur pour la méthode hybride. Dans le cas de l'estimation de la demande en transport, ce type de problème peut prendre des envergures importantes en fonction du territoire d'étude et, en conséquence, peut demander une capacité importante du solveur. L'intérêt de la résolution

par échelle de découpage est la possibilité de désagréger au fur et à mesure les déplacements estimés en gardant une qualité des résultats meilleure qu'avec la méthode classique.

Les résultats obtenus pour le cas de Montréal ont permis de dresser des profils d'utilisateur (par zone ou groupe de zone). L'analyse systématique des corrélations entre estimations et données socio-économiques permet de juger de la pertinence du modèle de manière intuitive. A posteriori, des résultats effectifs d'enquêtes ou des statistiques socio-économiques concernant le territoire pourraient être comparés à cette analyse et confirmer la performance du modèle hybride.

## CONCLUSION

Dans cette dernière partie du mémoire, des axes de réflexions concernant le modèle d'estimation de la demande TC sont présentés. Un retour sur la procédure classique de modélisation en premier lieu permettra de mettre à nouveau en valeur le besoin d'une méthode alternative tant à cause des limites de la PSC que des outils et des données disponibles. Par la suite, l'analyse critique du modèle hybride servira de nouveau point de départ pour l'optimisation de celui-ci. Le modèle hybride donne effectivement de meilleurs résultats que le précédent mais qui ne sont pas suffisamment performants pour une interprétation correcte du comportement des usagers du TC.

### Analyse critique du modèle classique

En plus des critiques que nous avons pu faire à chaque étape de la PSC, nous souhaitons revenir sur le niveau d'erreur du modèle. Les données à partir desquelles sont estimés les coefficients de la distribution sont issues de statistiques elles-mêmes et de fait, possèdent déjà un niveau d'erreur. Une valeur anormale dans les données peut engendrer la non mesurabilité d'un ou plusieurs coefficients ainsi qu'une interprétation erronée des résultats. Les estimations doivent être considérées avec leur marge d'erreur. L'application de la procédure classique pour une région urbaine où le planificateur aurait besoin d'un niveau de détail élevé n'est donc pas conseillée. Toute grande ville actuelle possède des particularités et des zones d'influence (centre ville, centre d'affaires, zone industrielle, zone résidentielle) qu'il est nécessaire d'isoler pour optimiser l'estimation de la demande en transport.

Par ailleurs, les outils du projet ont une capacité limitée et ne peuvent donc pas supporter un niveau de détail trop élevé tant pour le découpage que pour les variables socio-économiques. Pour pouvoir simuler la PSC complètement, il aurait été intéressant d'intégrer un outil plus performant au niveau de la résolution de problèmes sous contraintes. Cet outil serait un solveur de problème non linéaire pouvant estimer plus de 2000 coefficients dans un temps de résolution correct. Le solveur actuel nous impose

d'agréger considérablement le territoire d'étude, ce qui ne permet pas une interprétation aussi détaillée que le prévoient les objectifs du projet et l'obtention d'une matrice OD au niveau des SR est impossible.

La PSC est connue pour décomposer le choix d'un usager en étapes (les quatre étapes que nous avons définies précédemment). Ces séquences sont ordonnées d'une manière que l'individu ne suit pas forcément dans la vraie vie : en fonction de son quartier, choisir de se déplacer en TC puis sélectionner la zone de destination par rapport à la distance (dans notre cas). Cette approche particulière reste cohérente puisqu'il s'agit d'une attitude logique de l'usager mais qui ne s'applique pas de la même manière pour tous les motifs de déplacements. Nous avons supposé que l'abondance de variables socio-économiques et de population suffiraient à palier les lacunes du modèle (traduire les zones particulières comme le CV) mais les résultats n'ont pas réussi à expliquer la réalité.

Néanmoins, il reste que, dès le départ, la pertinence et la qualité des données est primordiale. Si une erreur quelconque est présente, la modélisation même avec des indices de performance élevés sera mauvaise et pourra conduire à des interprétations incorrectes de la demande. Il est donc important, pour chaque variable, de vérifier qu'il n'existe pas de valeur erronée (anormalement grande ou petite) et ce notamment pour les secteurs connus (centre-ville, zone industrielle...) ainsi que de vérifier la pertinence des variables clés (socio-économiques, population).

### **Analyse critique du modèle hybride**

L'objectif principal du modèle hybride est d'améliorer les estimations du modèle classique. Le second objectif est qu'il puisse donner des résultats suffisamment détaillés malgré les lacunes des outils à disposition. Effectivement, la méthode hybride peut être complètement codifiée sur notre outil solveur puisqu'elle s'adapte à la capacité de celui-ci. Néanmoins, un solveur plus puissant permettrait d'obtenir les estimations plus rapidement.

Lorsque le niveau d'agrégation du modèle est très fort et que la région est restreinte, les résultats sont généralement bons puisqu'il y a moins de variations particulières à expliquer et le modèle classique suffit. Cependant, l'interprétation et le niveau d'information des résultats n'est pas assez riche pour pouvoir évaluer le réseau TC et l'améliorer. À l'inverse, lorsque le modèle est totalement désagrégé, le modèle classique ne produit que des estimations erronées à cause de l'importance des variations et la multiplicité des comportements à modéliser. Ces résultats ne permettraient pas une analyse fiable du système de transport. Pour un niveau d'agrégation trop faible (ou à partir d'un certain niveau désagrégé), la méthode classique ne permet pas d'obtenir une erreur acceptable.

Notre étude ne nous a permis d'expérimenter que deux systèmes de découpage mais il aurait été intéressant de confronter la modélisation à différents niveaux d'agrégation du territoire. Les découpages administratifs proposés ne sont effectivement pas suffisamment flexibles dans le sens où l'on ne possède pas de détails sur les données et la définition des zones est normalisée. La modification de celles-ci donnerait une répartition inégale de la population (zones inégales) et donc des valeurs moins pertinentes pour l'étude.

La démarche agrégation-désagrégation de la méthode hybride permet d'obtenir par étape des résultats à différentes échelles de découpage (deux échelles dans notre cas), qui sont validés à partir des variables explicatives et de la matrice de référence. La qualité des données pour chaque système de zonage et leur cohérence sont donc primordiales pour la performance et la fiabilité du modèle. La vérification de la cohésion entre les types de données est nécessaire pour que les résultats soient aussi pertinents.

En addition des critiques de l'étape de distribution, le fait que le calage se base uniquement sur la matrice de référence pour ajuster les résultats peut représenter un handicap dans le sens où une telle matrice n'existe pas forcément (lorsqu'il s'agit de la première estimation de la demande pour le territoire) et peut aussi demander à être adaptée (issue de

statistiques). Il est donc essentiel que cette matrice soit fiable et que, s'il s'agit d'une matrice décrivant les déplacements dans le passé, les valeurs soient intégrées au modèle de manière adéquate.

L'étape de distribution ne distingue pas les modes TC qui pourraient influencer plus ou moins sur le comportement de l'usager : métro, bus, tramway, train. Par exemple, la présence du métro ou d'une ligne de bus à fréquence élevée proche du domicile motive les déplacements TC même si la destination du déplacement est éloignée. De même, la présence d'une ligne de bus à fréquence très faible incitera peu un habitant de banlieue à emprunter le TC même pour un court déplacement. Cette composante modale doit se retrouver dans les coefficients  $\alpha_i$  et  $\beta_j$ . Il serait intéressant de la mettre en valeur ou alors d'intégrer un nouveau coefficient qui traduise le mode empruntable dans la formule de distribution.

De même, dans l'étape de génération, les déplacements émis et attirés se basent sur les caractéristiques socio-économiques de la zone d'étude. Il serait alors aussi intéressant de développer les indicateurs de reproductibilité pour estimer des coefficients de distribution qui permettent, à partir des valeurs aux marges, d'effectuer une répartition des déplacements bruts qui seraient affinée par itération. De plus, au lieu des  $\alpha_i$  et  $\beta_j$ , l'utilisation de coefficients  $k_{ij}$  pour chaque paire OD ( $i,j$ ) serait plus pertinente si la capacité du solveur le permettait. Les résultats et leur interprétation seraient alors plus précis. Malheureusement, l'outil dont nous disposons est limité et ne peut pas supporter autant de contraintes pour cette étape.

Le nouvel axe de réflexion pour ce projet serait de mettre en place de nouvelles variables explicatives caractérisant le système de transport dans sa globalité et non seulement le TC : motorisation, existence d'une artère principale (autoroute), temps moyen d'accès à un arrêt (à pied), longueur des pistes cyclables; l'accès à des données sur les activités socio-

économiques serait aussi pertinent : nombre de commerces, présence d'un centre commercial, nombre d'écoles et places scolaires, présence de bus scolaire, nombre d'emplois par secteur. Ces données permettraient d'améliorer le modèle dans un premier temps mais aussi de mettre en valeur de nouvelles variables qui participent au choix de l'utilisateur d'emprunter le TC ou non.

Un autre axe de réflexion serait de considérer les déplacements multimodaux hors du centre ville. En effet, les stationnements incitatifs (« park and ride ») sont nombreux hors CV et permettent aux usagers de garer leur véhicule près d'un arrêt ou d'une gare principale généralement dans le but de se rendre au CV. Dans notre étude, nous n'avons considéré que les déplacements endogènes (au départ de la zone de résidence), cela peut signifier que l'utilisateur a pu emprunter son véhicule jusqu'à un stationnement de la zone avant de prendre le TC (il faudrait vérifier le déplacement précédant le déplacement TC) ou alors qu'il existe effectivement un arrêt ou une gare près de son domicile. Cette étude permettrait non seulement d'améliorer l'estimation des déplacements hors du CV mais aussi d'analyser l'usage et la performance des stationnements incitatifs (Les usagers à l'arrêt les utilisent-ils ? Quel pourcentage d'utilisateurs s'y gare pour se rendre hors CV ?)

La définition d'un coût généralisé moyen de transport serait également pertinente pour améliorer la fonction d'impédance pour la distribution. L'utilisation de la seule donnée de distance de parcours optimale en voiture n'est pas forcément pertinente pour distribuer des déplacements pour le transport en commun. En fonction de la structure du réseau TC (voies réservées, nombre d'arrêts sur une ligne, chemins particuliers des bus, utilisation du métro), ce type de distance peut ne pas être valide. Il serait donc intéressant d'estimer (et ce pour chaque territoire d'étude) les différentes composantes de ce coût telles que nous les avons présentées dans le chapitre 3.

Finalement, plusieurs moyens sont à entreprendre pour améliorer le modèle d'estimation. Les axes de réflexion que nous avons présentés ici sont essentiels dans le sens où ils aideront à l'interprétation des résultats, à la description plus pertinente du comportement des usagers. Ces problématiques impliqueraient les mêmes logiciels pour la génération et la distribution mais demanderaient l'intégration de nouveaux outils pour le calcul des variables explicatives.

## Synthèse

Les résultats de la méthode hybride pour le cas de Montréal ne paraissent pas performants. Malgré le fort niveau d'explicativité statistique de la régression (étape de génération), la distribution montre les lacunes des données utilisées pour estimer les déplacements (notamment pour la fonction d'impédance).

Néanmoins, l'utilisation de la méthode classique n'aurait pas donné de si bon résultats puisque le processus ne peut même pas être appliqué entièrement avec les outils du projet au niveau de découpage du territoire demandé. La PSC procède à l'estimation des déplacements des usagers d'un territoire donné de manière globale c'est à dire sans considérer les zones particulières ou les comportements singuliers de certains usagers.

Évidemment, la qualité des données et leur « interprétabilité » influent fortement le modèle résultant. Lorsque le comportement des usagers est trop complexe, le modèle qu'il soit classique ou non, ne peut expliquer qu'une partie des déplacements. Il est donc toujours important d'effectuer une étude préliminaire des corrélations et des variations de l'ensemble des variables et de les adapter (transformation en variable catégorique, segmentation) avant toute modélisation.

La vérification systématique de la cohésion des données entre les différentes échelles de découpage est aussi primordiale (pour le modèle hybride). L'estimation des déplacements à



partir de données issues de bases différentes et non fusionnées implique forcément des erreurs dans le modèle mais aussi dans l'interprétation des résultats.

Finalement, la nouvelle méthode que nous avons expérimentée est concluante car elle donne globalement de bons résultats par rapport à la méthode classique même avec peu de variables. La performance de cette méthode dépend fortement de la sélection des variables explicatives. Leur corrélation avec les déplacements doit être forte (colinéarité entre les variables) ou alors, lorsqu'aucune variable ne correspond, il faut effectuer plusieurs simulations avec différentes variables. Même si elle allonge l'étape de génération, la méthode reste simple et permet de bien différencier les variables significatives pour un ensemble précis de zones qui ne sont généralement pas les mêmes que pour la totalité des zones étudiées. Elle permet d'obtenir plus de détails et de pertinence dans le modèle.

## BIBLIOGRAPHIE

Beckmann, M., Mc Guire, C., & Winsten, C. (1956). *Studies in the economics of transportation*. Yale University Press.

Bonnel, P. (2001). *Prévision de la demande de transport*. Lyon: ENPC.

Bonnel, P. (2004, July). The estimation of aggregate modal split models. *World Conference of Transport Research, Istanbul* .

Campbell, M. (1950). *Route selection and traffic assignment: A compendium of correspondence relating to a suggested technique* . Highway Research Board.

Chapleau, R. (1986). Transit network analysis and evaluation with a totally disaggregate approach. *World Conference on Transportation Research* .

Chapleau, R., & Trépanier, M. (2001). Analyse orientée-objet et totalement désagrégée des données d'enquêtes ménages origine-destination. *Canadian Journal of Civil Engineering* 28

Chapleau, Robert; Lavigneur, Pierre. (1991, Septembre). Transport en commun et tendances socio-démographiques: situation québécoise. *Routes et Transports* , pp. 6-11.

Chu, X. (2004). *Ridership Models at the stop level*. National Center for Transit Research, University of Florida.

Domencich, T., & McFadden, D. (1975). *Urban Travel Demand: A Behaviour Analysis*. North Holland Publishing Co.

Douglas, A., & Lewis, R. (1971). *Trip Generation Techniques*.


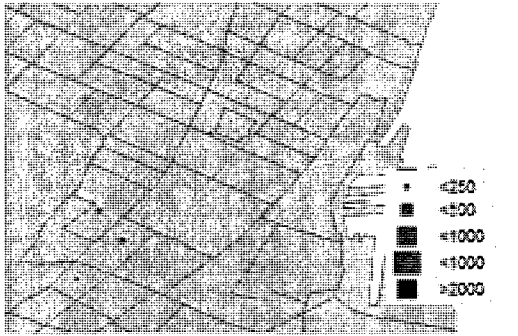
- Frank, L., & Pivo, G. (1994). Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, and Walking. *Transportation Research Record*.
- Furness, K. P. (1965). Time function iteration. *Traffic Engineering Control* 7.
- G. Vuk, T. J. (2007). Observed Modelled behavioural changes caused by the Copenhagen Metro. *Transport Science and Technology*.
- Hägerstrand, T. (1952). *The Propagation of innovation waves*. Lund.
- Hitchcock, F. (1941). The distribution of a product from several sources to numerous facilities. *Journal of Mathematical Physics*, vol. 20, pp. 224-230.
- Kanafani, A. (1983). *Transportation demand analysis*. McGraw-Hill College.
- Kittelson & Associates, Inc. (2001). *TLOS Software Users Guide*.
- Kohl, J. (1841,1850). *Ansiedlungen der Menschen in ihrer Abhängigkeit von der Gestalt der Erdoberfläche*. (T. Peucker, Trad.)
- Lill, E. (1891, traduction 1969). *The Law of Travel and Its Application to Rail Traffic*.
- Manheim, M. (1979). *Fundamentals of Transportation Systems Analysis; Volume 1: Basic Concepts*. MIT Press.
- Meyer, M., & Miller, E. (1984). *Urban Transportation Planning: a Decision-Oriented Approach*. McGraw-Hill.
- Mitchell, R., & Rapkin, C. (1954). *Urban traffic: A function of land use*. New York: Columbia University Press.
- Oi, W., & Shuldiner, P. (1962). *An Analysis of urban travel demands*. Evanston, Ill.: Northwestern Univ. press.

- Perone, S., Carr, T., & Upton, D. (2006). Forecasting Traffic Impacts on a Citywide Street Network in the Absence of a Travel Demand Model. *Transportation Research Record* .
- Ravenstein. (1885,1889). *The laws of migration* (Vol. 52). Journal of the Royal Statistical.
- Reilly, W. (1931). *The Law of Retail Gravitation*.
- Salter, R. J. (1974). *Highway traffic analysis and design*.
- Stewart, J. (1948). Demographic Gravitation: Evidence and Applications. *Sociometry*, Vol. 11, No. 1-2 , pp. 31-58.
- Stouffer. (1940). Intervening opportunities : a theory relating mobility and distance . *American Sociological Review* , pp. 845-867.
- V. Arasan, M. W. (1996, Septembre/Octobre). Modelling of Stratified Urban Trip Distribution. *Journal of Transportation Engineering* .
- V.L., B. (2008). Accessibility-Based Approach to Travel Demand Forecasting: New Alternative to Four-Step and Activity-Based. *Transportation Research Board* .
- Wardrop, J. (1952). Some Theoretical Aspects of Road Traffic Research. Dans *Proceedings of the Institute of Civil Engineers, part II, Vol 1* (pp. 325-378). London.
- Wilson, A. (1974). *Urban and Regional models in geography and planning*. London.
- Zipf, G. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge: Addison-Wesley.

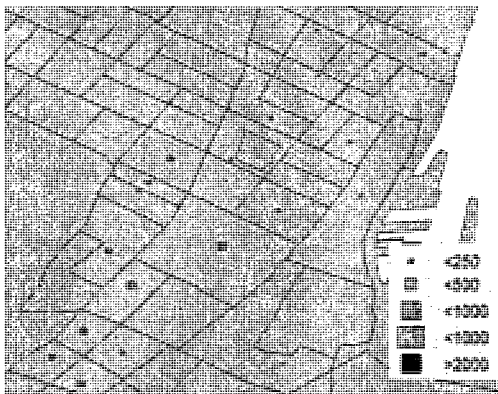

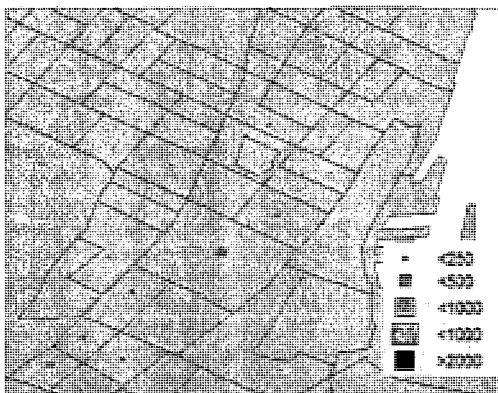
**ANNEXE A – Étude des déplacements par motif pour le cas de la ville de Montréal**

L'étude préliminaire des données de l'enquête OD 2003 pour la ville de Montréal nous a permis d'étudier le comportement des usagers du TC en terme de quantité de déplacements par rapport au motif.



**Tableau A.1: Comparaison du nombre de déplacements entrants à l'intérieur des zones du CV, par motif de déplacement**

	Motif travail INTRA CV	On se rend compte que les déplacements se concentrent sur la partie centrale ouest du CV.
	Motif études INTRA CV	Les déplacements pour motif études sont inférieurs au motif travail. Les SR concernés sont à la périphérie du CV.



**Tableau A.1: Comparaison du nombre de déplacements entrants à l'intérieur des zones du CV, par motif de déplacement (suite et fin)**

	Motif loisirs INTRA CV	Le motif loisirs est celui qui concerne le plus d'utilisateurs. Également ici, les déplacements se font surtout dans la partie centrale ouest du CV.
	Motif magasinage INTRA CV	Le motif magasinage ne concerne essentiellement que deux SR. Il serait intéressant d'y observer les commerces de ces SR.
	Motif autre INTRA CV	Pour les autres motifs, la trame de la carte est similaire à celle du motif travail.

**Tableau A.2: Comparaison du nombre de déplacements entrants à l'intérieur des zones hors CV, par motif de déplacement**


	<p><b>Motif travail HORS CV</b></p> <p>On remarque essentiellement une concentration des déplacements au niveau de la zone Plateau-Rosemont Petite Patrie-Villeray.</p>
	<p><b>Motif études HORS CV</b></p> <p>Ici, la répartition des déplacements est disparate, elle doit correspondre aux écoles et université et zones résidentielles de l'île.</p>

**Tableau A.2: Comparaison du nombre de déplacements entrants à l'intérieur des zones hors CV, par motif de déplacement (suite)**

	<p><b>Motif loisirs HORS CV</b></p> <p>Les déplacements pour motif loisirs sont plus marqués dans les SR et les zones correspondantes sont clairement bien délimitées : l'ouest de l'île, Verdun-Lasalle, Côte des Neiges-Mont Royal, Plateau-Rosemont, Montréal Nord.</p>
	<p><b>Motif magasinage HORS CV</b></p> <p>Le motif magasinage est moins recensé et se limite à la partie centrale de l'île.</p>



**Tableau A.2: Comparaison du nombre de déplacements entrants à l'intérieur des zones hors CV, par motif de déplacement (suite et fin)**

	<p>Motif autre HORS CV</p> <p>Ici aussi le dernier motif est peu interprétable. On reconnaît essentiellement les quartiers Montréal Est, Côte des neiges et Plateau.</p>
---	--

## ANNEXE B – Exemple de matrice OD

Nous proposons ici une partie de la matrice OD pour les secteurs municipaux 101 à 112 (où 101 et 102 constituent le CV). Elle représente le taux de déplacements réalisé entre deux SM.

**Tableau B.1: Exemple de matrice OD pour les secteurs municipaux 101 à 112 du cas de Montréal**

SM	101	102	103	104	105	106	107	108	109	110	111	112
101	0,105	0,001	7E-09	2E-09	8E-06	2E-08	7E-09	2E-06	8E-10	2E-09	2E-08	2E-07
102	0,043	0,072	0	0	2E-04	4E-07	2E-07	4E-05	2E-08	4E-08	4E-07	6E-06
103	0,078	0,024	0,055	6E-06	0,007	8E-06	2E-06	0,001	6E-07	1E-06	1E-05	2E-4
104	0,074	0,033	3E-05	0,051	0,066	1E-05	7E-06	0,005	2E-06	2E-06	1E-05	3E-4
105	1E-4	3E-05	2E-08	1E-08	0,126	3E-08	2E-08	5E-06	2E-09	2E-09	1E-08	3E-07
106	0,099	0,059	3E-06	2E-06	0,011	0,037	2E-05	0,003	2E-06	7E-06	4E-05	4E-4
107	0,080	0,04	2E-06	1E-06	0,019	7E-05	0,083	0,018	9E-06	1E-05	2E-05	5E-4
108	5E-05	2E-05	4E-09	3E-09	1E-05	2E-08	6E-08	0,083	9E-09	6E-09	2E-08	6E-07
109	0,045	0,020	4E-06	2E-06	0,009	4E-05	8E-05	0,019	0,061	5E-05	6E-05	0,002
110	0,102	0,049	8E-06	1E-06	0,019	8E-05	4E-05	0,007	3E-05	0,06	3E-4	0,003
111	0,056	0,035	3E-06	1E-06	0,004	2E-05	5E-06	0,001	4E-06	1E-05	0,076	0,001
112	2E-4	8E-05	2E-08	5E-09	2E-05	7E-08	4E-08	2E-05	3E-08	4E-08	3E-07	0,078